GOOD TIMES... BAD TIMES

How positive and negative associations are hard-wired in the brain **PAGE 675**

MIDDLE EAST

ARAB SPRING AFTERMATH
Uncertainty reigns in science and education
PAGE 604

HEALTH SCIENCE

TRIAL SEPARATION
N-of-1 tests needed if medicine is to get personal
PAGE 609

A GOOD READ

BOOK REVIEWS SUPPLEMENT
Climate economics, junk DNA and more
PAGE 614

Thank you for sharing

Initiatives to make genetic and medical data publicly available could improve diagnostics — but they lose value if they do not share with other projects.

The modern world is all about sharing, driven by the borderless flow of information through the Internet. Pictures, articles, jokes, links, ideas, criticism — information has never been so free to move. And from open access to giant web-based data repositories, science in 2015 is increasingly based on shared knowledge and expertise.

Sharing should be equal, but some is more equal than others. The principals behind one genetic data-sharing project unveiled last week have described their initiative as a model of "scientific openness" that offers "broader access" to genetic data. Indeed, the name of the project — BRCA Share — trades on the idea of data freedom. The initiative focuses on clinical data concerning mutations in the genes *BRCA1* and *BRCA2*, which increase risk of breast and ovarian cancer.

In truth, it creates more of a walled garden of genetic data than an open field. That runs contrary to crucial ongoing efforts to amass large amounts of linked genetic and medical data to help scientists and doctors to improve interpretation of genetic test results.

The commercial market for genetic testing for breast cancer is growing. On 21 April, for instance, a firm called Color Genomics in Burlingame, California, promised to offer women *BRCA* gene testing — which normally costs thousands of dollars — for as little as US$249. If the companies that offer such testing share their data, stripped of identifying information, with researchers, it could aid efforts to understand how all of the thousands of possible *BRCA* mutations affect cancer risk. The more data researchers can gather, the more they can determine whether 'variants of unknown significance' — genetic differences whose health effects are unknown — are benign or pose risks.

BRCA Share is a partnership between Quest Diagnostics of Madison, New Jersey, and INSERM, the French National Institute of Health and Medical Research. Testing labs that sign up will get access to each other's *BRCA* data and an INSERM repository of information about the genes. The project will run functional studies on the effects of mutations.

But it will not share data with similar efforts such as ClinVar, a US National Institutes of Health-funded initiative that is making linked genetic and medical data publicly available for all. Quest says that BRCA Share cannot contribute to ClinVar because its data are structured differently. But this seems a weak excuse. Researchers say that increasingly, where there is a will to share, a technological way can be found. Because of the restrictions, several labs have declined to join BRCA Share.

Quest says that participants in BRCA Share will be allowed to share their own data with ClinVar. But the first to join the initiative — diagnostics company LabCorp of Burlington, North Carolina — has so far chosen not to; it has not deposited any *BRCA* data in ClinVar. Quest says that it will share BRCA and other genetic data with the Human Variome Project and the Leiden Open Variation Database, based in the Netherlands, which has an agreement to share data with ClinVar, but the agreement between Quest and Leiden has not yet been signed.

The episode showcases an uncomfortable truth about personalized medicine: everyone agrees that large data sets are crucial, and everyone is racing to collect them. The larger the data set, the more useful. The most useful of all would be one huge database containing all available data. But even though all parties recognize the value of it, many are choosing not to share, and this holds back medical progress.

BRCA Share is only one example; there are many others. The company 23andMe in Mountain View, California, for instance, has collected genetic data on 900,000 people. It shares aggregate statistics with outside researchers and has published or contributed to 32 papers in the past 5 years. But to protect users' privacy, it does not routinely share linked genetic and medical data.

> *"Even though all parties recognize the value of it, many are choosing not to share."*

The field is young and will take time to settle. Already, dissatisfaction with limits on sharing is driving new initiatives, such as Genes for Good at the University of Michigan in Ann Arbor, which is enlisting Facebook users for a genetic study (see page 597), and the online community Open Humans, which is helping participants in genetic studies to share their data with other researchers.

Awareness of the issue among consumers is also driving companies to share more widely. Color Genomics, for instance, promises to share its *BRCA* data with ClinVar, as do many other companies.

The imperative to share data remains an esoteric issue for much of the public, and one that is easily obscured. Companies or researchers who talk the talk of sharing but do not actually walk the walk should be challenged. Data sharing is too important to be turned into meaningless marketing speak. ∎

A hard sell

Scientists must stand up for marine parks if the value of the seas is to be recognized globally.

Studland Bay is an unlikely battleground. The sandy shore, part of England's southern coastline, is both a beauty spot and the site of "the most popular naturist beach in Britain". More importantly for *Nature*-ists, the seagrass that thrives in the bay's shallow waters is home to rare sea-horse populations. But keeping it that way is a growing challenge, because Studland is also a playground for the wealthy and powerful, including some yacht-club members who like to anchor their boats there, to the possible detriment of the natural habitat.

Scientists can make the case for conservation, but the value of marine sites such as Studland is hard to sell. It is not the Great Barrier Reef with its miles of coral, or the Galapagos Islands and their sharks, turtles and marine iguanas. It is not even Lundy — the island at the

centre of the first UK marine park, home to seals and puffins.

Last week, the UK government ended a public consultation on the latest round of its attempts to manage conflicts between conservation and other maritime interests. Officials plan to create dozens of marine conservation zones (MCZs) around the UK coast, to introduce an extra level of protection for worthy sites. The move is part of a global effort to extend conservation measures on land to the coast and open seas.

But, as on land, conservation of marine areas is not popular with everyone, and the ambitious UK plans are in danger of running aground. Already, proposals for the Studland MCZ have been dropped for now, and plans for others are in danger. In fact, marine experts who originally identified 127 sites for MCZs around the British coastline now fear that, in the worst case, barely more than half will be realized.

The UK process sets up these zones in three waves. The first saw 31 proposed MCZs reduced to 27, which were designated in November 2013. A second tranche of designations initially featured 37 sites, but was whittled down to 23 even before the public consultation closed last week. (Studland Bay MCZ was one of the 14 that were chopped.) The third wave is meant to be set up before 2016.

The United Kingdom is squaring up for a general election on 7 May, and even long-serving politicians are nervous about keeping their seats. Marine conservation is low on the agenda — especially the sort that annoys boat-owning voters. But the next government will have to take some tough decisions on marine protection. Because of the way the first two tranches of MCZ designation have progressed, the third will be left with difficult decisions on controversial sites that have been kicked into the long (sea)grass by the earlier rounds. These include not only Studland but also the important deep-water mud habitats of the Celtic Deep, and another sea-horse habitat off the Isle of Wight.

Crucially for the United Kingdom, failures at home are undermining its potential to show global leadership on marine conservation. Huge marine parks have been designated around UK territories such as the Chagos Islands in the Indian Ocean, and another was announced on 18 March for the Pitcairn Islands in the south Pacific. Significant questions remain over enforcement and monitoring of these overseas reserves, and must be solved if they are not to become what scientists call 'paper parks' — just a line on a map rather than something that does any good.

> *"The UK government deserves credit for establishing the parks at all."*

But the UK government deserves credit for establishing these parks at all. This is despite a recent setback for the Chagos reserve — a judge ruled last month that the park had been improperly declared by Britain because Mauritius, which has fishing rights in the archipelago, had not been fully consulted.

Later this year, the finalized United Nations Sustainable Development Goals are likely to include a commitment to protecting the oceans. Making good on this and on existing promises will require governments to disappoint some powerful vested interests that would prefer to keep access to marine areas unrestricted. And if rich nations wish to push the rest of the world to protect its marine wonders, they must first get their own houses in order. Scientists can help by promoting the value of places that lack the brand recognition of the Great Barrier Reef. If the world's seas are really going to be protected, mud and seagrass will have to be considered alongside tropical reefs. ∎

More on unicorns

A newly discovered tiny dinosaur sported an intriguing structural accessory.

How do you go about hunting for unicorns? More specifically, let's say that you have found your unicorn, but then what? How would you know that it was a unicorn, rather than something else? How diagnostic, say, is that horn? Might there not be other creatures, as yet unimagined, that have unicorn-like horns but differ from unicorns (and everything else) in all kinds of unusual ways? Narwhals, for example, exist — and their unicorn-like horn is the only feature they have in common with unicorns. So what else might lurk out there in the great unknown? You know that unicorns don't number among the domestic beasts, says Jorge Luis Borges in his essay 'Kafka and his precursors', but, without seeing one, how do you know what a unicorn is "like"?

Palaeontology is always about hunting for unicorns. Because fossils sample diversity from the past that might not exist today, there is always a chance that one will dig up something that defies categorization. Such is the case for the tiny dinosaur described online in *Nature* by Xing Xu and colleagues (see go.nature.com/jsxjxv), and discussed further by Kevin Padian (go.nature.com/s6g2aw).

The dinosaur is about as far from *Brontosaurus* and other behemoths familiar to the museum visitor as might be imagined. It belongs to a little-known group of creatures whose body size has hitherto scaled inversely with the length of their names. It is only the third known example of a (deep breath) scansoriopterygid, the other two bearing the unwieldy handles *Epidendrosaurus* and *Epidexipteryx*. Xu and colleagues buck the trend by calling their creature *Yi*. Or, in full, *Yi qi*. This must be the shortest dinosaur name ever, and it is commensurate with the tiny size of scansoriopterygids, which could have looked thrushes or starlings in the eye. Although scansoriopterygids had feathers and cluster phylogenetically round the ancestry of birds and other feathered dinosaurs, they do not seem to have had flight feathers on their disproportionately long forelimbs. Reconstructions make them out to be rather like feathered lemurs, scampering along branches, perhaps wheedling insects out of crevices with their long, clawed fingers.

Except that *Yi qi* is different. Attached to each wrist is a strut, made of bone or calcified cartilage, which cannot simply be homologized with regular wrist or hand bones. The strut seems to be a new structure made from an accessory wrist bone, possibly a sesamoid — the kind of bone usually embedded in a tendon or muscle. The 'thumb' of the giant panda is made of just such a bone. The sesamoid of *Yi qi* (if that's what it is) is much larger, however, in relation to the animal as a whole — equal in length to the bones of the forearm. It had to be there for a reason, but what was it?

It is here that we enter unicorn territory — for no dinosaur, however unusual, has been found with anything like this feature. The authors are appropriately cautious, therefore, in their interpretation. They point to the hint of a suggestion that some soft tissue, preserved alongside these curious elements, represents what might have been a membrane that the sesamoid bone supported. From that, they suggest that *Yi qi* had membranous wings and might have glided from branch to branch, in much the same way as various tree-living mammals and reptiles do today. But it was probably not capable of powered flight as birds and bats are — and as were, presumably, the extinct pterosaurs, which were (one must stress) only distant relatives of dinosaurs and birds.

When, in the mid-1990s, the first dinosaurs with preserved feathers came to light, there was a great deal of celebration. However, evidence for the bird-like nature of dinosaurs had been accumulating for a while, so for many (though not all) people, feathered dinosaurs were a vindication rather than a challenge. *Yi qi*, by contrast, is something else. Here we have a feathered dinosaur and a close relative of birds that seems to have essayed an entirely different experiment in aerial locomotion. For a feathered dinosaur to have traded feathers for a membrane in an aerofoil is something nobody could have predicted. Whether or not it is a unicorn has yet to be determined. ∎

↻ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

China's scientific progress hinges on access to data

It is getting harder for scientists in China to obtain the high-quality public data that they need for important research studies, says **Zheng Wan**.

The era of big data has seen the development of bioinformatics databases, data sharing and increased access to technical resources. Until recently, this trend towards open science was helping scientists in China to compete on the world stage. However, growing constraints on public-data exploration and sharing are being felt throughout Chinese society. If not addressed, they will slow scientific research and innovation.

It is hard — and getting harder — for Chinese scientists to access high-quality domestic data. Most of the public data are held by government departments, some of which are strengthening their monopoly and making it harder for researchers to access the information. This affects researchers in the humanities and social sciences especially, but also extends to fields such as environmental science and public health, because the data involved can be politically sensitive. At conferences, I hear numerous complaints from colleagues about how hard it is to extract routine figures such as air-pollutant levels from the authorities, for example.

Even when data are published, some are likely to be of poor quality because they have not been collected properly. The most notable example is the controversy on China's gross domestic product (GDP). There is a significant — and widening — difference between the official national estimate and the total calculated by adding up the GDPs of each of China's 31 province-level divisions. The National Bureau of Statistics in Beijing admits that different data-collection methods are used at the provincial level, and is trying to harmonize them. So far, progress has not been encouraging.

Public data sharing has been turned into a profit-making scheme. It would be useful, for example, to compile data on pollution from road vehicles in China. Done properly, this would require access to detailed records on the number of each type of vehicle licensed, road congestion, detailed engine parameters and fuel standards. Research institutions struggle to get even basic figures on vehicle ownership from public agencies, so they must use less rigorous — and often misleading — sales data that industry groups collect from manufacturers. Ironically, the same wealthy automobile manufacturers that inflate their own numbers can get objective, reliable data about their competitors' sales by buying them from special channels that are linked with some government departments — at a price that public institutions and scientists cannot afford.

In such an environment, it is no surprise that some research teams in China do not want to publish their own data. Ownership of data represents intangible capital that gives scientists a competitive advantage in some academic fields. My own research group receives many requests for maritime data, such

> ## SECURE NETWORKS ARE CRUCIAL FOR NATIONAL SECURITY, BUT GOOD DATA ARE THE BACKBONE OF SCIENTIFIC PROGRESS.

as port statistics and fleet information, that we have compiled, but we are reluctant to share the information. The workload and cost of sourcing and sorting scattered data sets into a usable form are enormous. If we keep these data exclusive, we can use them to develop research papers. If it were easier to access good data from other sources, we would be more comfortable with giving our own away.

Open access to and improved quality of public data can promote transparency in government affairs. Despite the slow progress, there are positive examples of improved transparency and how it has benefited Chinese society. A notable one is the full disclosure of air-pollution data, which started in 2014 as a result of mounting public pressure on environment bureaus. Before this, only sketchy data were published on a daily basis. These data are now updated hourly and are widely shared between government agencies in China. The data allow the health bureau to send alerts to the general public, the education bureau to decide whether classes should be suspended during smoggy days and the transportation bureau to adjust its vehicle-restriction rules. Perhaps the most important effect of the full disclosure is heightened public awareness of the worsening pollution crisis. Unfortunately, not all data that interest scientists also interest the general public. There is still a long way to go in achieving full transparency and increasing the availability of public data.

The restrictions are not confined to information generated and held inside China. Foreign academic resources can also be technically demanding to access. Several information-management bureaus have set up digital roadblocks to filter supposedly harmful information. My life as a working scientist in China is affected. Reliable searches of the academic literature are near impossible. With no access to Google Scholar — which I prefer over other search engines, because it combines books, papers, theses, patents and technical reports — I have to keep track of trends by individually searching the databases operated by publishers that are still accessible.

Secure networks are crucial for national security, but good data are the backbone of scientific progress and economic development. Resource sharing and public access to trustworthy information underpin economic and social well-being. In China, resolving these conflicts will entail a comprehensive study, so that we can establish a highly efficient and reasonable data-management mechanism that benefits all. More immediately, researchers should be granted greater access, especially to public data and to academic search engines. ∎

Zheng Wan *is a research scientist at the College of Transport and Communications, Shanghai Maritime University, China.*
e-mail: mrwan@ucdavis.edu

RESEARCH HIGHLIGHTS
Selections from the scientific literature

Light direct from an alien world

Astronomers have spotted light reflected off a planet orbiting a distant sun, by teasing it out from the background starlight. The discovery allows direct calculations of the mass and other properties of the exoplanet, rather than inferring them using other methods.

Jorge Martins of the University of Porto in Portugal and his team used the HARPS instrument at the European Southern Observatory's La Silla facility in Chile to study 51 Pegasi b, which was the first ever exoplanet found around a Sun-like star. The team subtracted the starlight, leaving only the faint planetary spectrum. From this, they calculated the planet's mass to be about half that of Jupiter.

Detecting this reflected light from other exoplanets could allow astronomers to determine more characteristics than other techniques can.
Astron. Astrophys. 576, **A134** (2015)

Bacteria swim to form crystals

Fast-swimming bacteria that live at the bottom of salt marshes spontaneously come together to form organized crystalline structures that move through water.

The bacterium *Thiovulum majus* is a large, round cell with hundreds of small flagella, or tail-like structures, that spin to propel the organism. Alexander Petroff of the Rockefeller University in New York and his colleagues observed the behaviour of individually spinning cells under a microscope and found that each created tornado-like flows

How grebes walk on water

The heaviest animals known to run on water pull off the feat by using quick strides and large feet that slap the surface.

Western and Clark's grebes (*Aechmophorus occidentalis* and *Aechmophorus clarkii*; pictured) run as far as 20 metres on water and for up to 7 seconds during a mating ritual, and are among only a few animals with this ability. Glenna Clifton at Harvard University's Concord Field Station in Bedford, Massachusetts, and her team analysed high-speed video of wild birds performing the dance and studied models of the grebe foot in the laboratory.

They found that the birds stay above the water by having a fast stride rate of up to 20 steps a second, as well as wide, flat feet that slap the water surface with enough force to support up to 55% of the bird's weight. The animal's feet are also shaped to reduce drag.
J. Exp. Biol. 218, **1235–1243 (2015)**

in the liquid around them. These forces cause the cells to self-assemble into groups of between 10 and 1,000 cells, arranged in a two-dimensional crystalline hexagonal shape that rotates collectively.
Phys. Rev. Lett. 114, **158102 (2015)**

Groundwater under Antarctica

A groundwater network found beneath an Antarctic valley could support microbial life.

Jill Mikucki of the University of Tennessee in Knoxville, Slawek Tulaczyk at the University of California, Santa Cruz, and their colleagues studied Taylor Valley using an airborne instrument to measure underground electrical resistivity, which increases as water in soil freezes. The team discovered two systems of briny groundwater, one of which connects to the Ross Sea.

This water could be home to microbes, the authors suggest, because subsurface temperatures at boreholes were within the range that can support microbial life (–3 °C to –20 °C). And an outflow of iron-containing subsurface saltwater from the Taylor Glacier is known to contain microorganisms.
Nature Commun 6, **6831 (2015)**

Brain waves go far in tinnitus

Many people hear ringing in the ears, known as tinnitus, and researchers now think it could involve abnormally linked brain waves that extend across a large part of the brain.

William Sedley at Newcastle University, UK, Phillip Gander

at the University of Iowa in Iowa City and their team recorded brain activity in an awake 50-year-old man with tinnitus, who had electrodes implanted to study his epilepsy.

The researchers used sounds to temporarily suppress the loudness of the tinnitus, and looked for shifts in brain-wave patterns as the tinnitus intensity changed. They found that low-frequency brain waves previously linked to tinnitus spread throughout a larger proportion of the auditory cortex than was thought. Mid- and high-frequency oscillations affected brain regions involved in auditory memory and perception.

The three sets of brain waves seem to interact to create a widespread system of abnormal activity.
Curr. Biol. **http://doi.org/338 (2015)**

METROLOGY

Atomic clock smashes records

The world's best atomic clock has become even more accurate, neither gaining nor losing a second over 15 billion years — longer than the age of the Universe.

The clock keeps time using 2,000 ultracold strontium atoms trapped in a laser lattice. Its pendulum is a laser that has a frequency tuned to resonate with that of the photons that are released and absorbed by the atoms as they switch between two energy states. By using a more stable laser and by reducing the effects of environmental perturbations, Jun Ye at the JILA institute in Boulder, Colorado, and his team made the 'ticks' of the timepiece more uniform. This increased the clock's accuracy by three times compared with the previous record.

The result brings strontium-based optical clocks closer to replacing the current standard for time measurement, the caesium fountain clock, which makes lower-frequency atomic transitions.
Nature Commun. **6, 6896 (2015)**

NAT. HIST. MUS. LONDON/SPL

CLIMATE CHANGE

Weather extremes linked to warming

Global warming is responsible for three out of four hot temperature extremes around the world, and almost one-fifth of heavy precipitation events over land.

Erich Fischer and Reto Knutti at the Swiss Federal Institute of Technology in Zurich used climate simulations to compare the frequencies of daily hot and wet extremes over the past century with those projected for the twenty-first century under a high-emissions scenario.

They found that with 2 °C of warming above pre-industrial levels, extreme events that were once expected to occur every 30 years or so might happen every 10–20 years.
Nature Clim. Change **http://dx.doi.org/10.1038/nclimate2617 (2015)**

PHYSIOLOGY

Colour tunes the body clock

The changing brightness of sunlight during dawn and dusk is known to adjust circadian rhythms, but researchers now show that the shifting colour of the light does the same thing in mice.

Timothy Brown, Robert Lucas and their co-workers at the University of Manchester, UK, studied mice that were genetically engineered so researchers could stimulate the two colour-detecting pigments in the mouse retina one at a time. When they recorded the firing of neurons in the brain's clock centre, they found that this activity was in sync with changes in the colour of light between blue and yellow.

Mice exposed to lighting conditions that mimic twilight, including colour changes, showed differences in the timing of their daily body-temperature fluctuations, compared with animals exposed to light that shifted only in brightness. The authors

suggest that the mammalian circadian clock uses colour to gauge the Sun's position during twilight.
PLoS Biol. **13, e1002127 (2015)**

PALAEOGENETICS

Genomes reveal mammoth history

Genome sequences from two woolly mammoths provide a rare look at the genetic events leading up to extinction.

Eleftheria Palkopoulou and Love Dalén at the Swedish Museum of Natural History in Stockholm sequenced

the genomes of a woolly mammoth (*Mammuthus primigenius*; artist's impression **pictured**) that roamed Eurasia roughly 45,000 years ago, and of one that lived on a remote island some 4,000 years ago, just before the animals went extinct. Using the genomes to infer population trends, the team found that mammoths declined around 300,000 years ago, then recovered until about 12,000 years ago when mainland populations went extinct, leaving just a few hundred island-dwelling animals.

The island mammoth showed decreased genetic diversity, which could have contributed to extinction, the team says, adding that this kind of genetic analysis could help conservation biologists to assess how much diversity has been lost from endangered species today.
Curr. Biol. **http://doi.org/34d (2015)**

↻ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

Embryo editing

The first ever report of genome editing in human embryos sparked debate on 22 April. Scientists in China used a technique called CRISPR/Cas9 to edit the gene responsible for the blood disorder β-thalassaemia (P. Liang *et al. Protein Cell* http://doi.org/34q; 2015). The work has split opinion over the ethics of such research. A team led by Junjiu Huang, a gene-function researcher at Sun Yat-sen University in Guangzhou, applied the technique to 86 embryos, but the faulty gene was replaced in only 4. Many embryos were found to have 'off-target', or unintended, mutations. The team used non-viable embryos, which could not have resulted in live births because they were produced by eggs fertilized by two sperm. See page 593 for more.

Malaria vaccine

A candidate malaria vaccine known as RTS,S/AS01, which has been in development for nearly 30 years, has failed to prevent deaths in a trial of more than 14,000 toddlers and infants. The results, reported on 24 April (RTS,S Clinical Trials Partnership *Lancet* http://doi.org/34s; 2015), showed that the vaccine cut the rate of uncomplicated malaria by 26% in babies aged 6–12 weeks who received a booster dose, and by 18% in those who did not. It had no effect on severe malaria. Regulators will now decide whether and how the vaccine should be used.

Europe's GM plan

Individual member states of the European Union (EU) would be able to ban genetically modified (GM) crops otherwise approved for use across the union, under plans by the European Commission. If adopted, legislation proposed on 22 April would allow countries to opt out of growing approved GM crops and to ban their importation for human food or animal feed. The EU pledged the reform in December 2014 (see *Nature* http://doi.org/xmq; 2014). Some researchers hope that it will allow such crops to be approved after years of stalling and debate between pro- and anti-GM nations.

Thai fishing notice

Fish imports from Thailand could be banned in much of Europe in a row over illegal fishing. The European Commission issued a formal warning to Thailand on 21 April, saying that the country was not doing enough to stop illegal fishing. Thailand has six months to improve monitoring, control and sanctioning systems, or it risks a trade ban.

US energy review

The United States must increase investment in its ageing energy infrastructure to maintain safety, reduce greenhouse-gas emissions and prepare for global warming, the US Department of Energy urged in its Quadrennial Energy Review on 21 April. The report recommends ten-year investments of between US$2.5 billion and $3.5 billion to replace and maintain oil and gas pipelines; $3 billion to $5 billion for programmes to ensure energy reliability; and $3.5 billion to modernize the electricity grid. The White House also announced a new partnership with major energy suppliers, intended to bolster energy infrastructure against extreme weather and climate change.

Hawaiian protest

Protests over the Thirty Meter Telescope (TMT) escalated across Hawaii last week, as Native Hawaiian demonstrators spoke out

NAVESH CHITRAKAR/REUTERS

Death and destruction in Nepal

Widespread destruction and loss of life came to Nepal as a magnitude-7.8 earthquake hit just 80 kilometres northwest of the capital, Kathmandu, on 25 April. As *Nature* went to press, more than 4,000 people had been reported dead, although it is feared that the final death toll will be much higher. The quake happened along one of the world's biggest geological collision zones, in the Himalayas. Many older buildings with unreinforced masonry collapsed; temples and cultural heritage sites, including those in Bhaktapur (pictured), have been destroyed, says the United Nations Educational, Scientific and Cultural Organization.

against the mega-telescope planned for construction on the sacred mountain of Mauna Kea. The project has legal authority to proceed, but has held off since early April because of the controversy. On 21 April, protestors marched to the offices of governor David Ige in Honolulu, asking for the TMT to be halted permanently. They say it is a desecration of native land. See go.nature.com/hcb6yy for more.

Transatlantic trek

The world's largest liquid-argon neutrino detector has found a new home at Fermilab near Chicago, Illinois, an international collaboration announced on 22 April. The ICARUS detector, which weighs 760 tonnes when full, will journey to the United States from the CERN particle-physics lab near Geneva, Switzerland, where it is being upgraded. Starting in 2018, ICARUS and two similar detectors will hunt for 'sterile' neutrinos, candidates for dark matter. Between 2010 and 2014, it was at the Gran Sasso underground lab in Italy, where it monitored a now-decommissioned beam of neutrinos shot through the Earth's crust from CERN.

Chilean volcano

The Calbuco volcano in southern Chile erupted on 22 April for the first time since 1972. The blast sent an ash plume towering more than 15 kilometres above the cities of Puerto Montt and Puerto Varas (**pictured**). Ash also fell across much of the neighbouring countryside. Calbuco is one of the most active volcanoes in the southern Andes, and in 1893–94 caused one of the biggest eruptions seen in Chile in recent history.

Head-injury ruling

A US judge has approved a 65-year settlement to resolve thousands of concussion-related lawsuits brought against the National Football League by retired American footballers. The 22 April decision is expected to cost the league about US$1 billion in total, with awards averaging $190,000 for conditions including neuromuscular diseases, cognitive impairment and certain cases of chronic traumatic encephalopathy (CTE), a degenerative brain disorder diagnosed only at autopsy (see *Nature* **518,** 466–467; 2015). The settlement excludes mood disturbances sometimes attributed to CTE, because of the difficulty of ruling out other causes.

Whistleblower win

The US Environmental Protection Agency (EPA) illegally retaliated against a chemist, Cate Jenkins, who exposed a cover-up of pollution dangers at the World Trade Center site after the terrorist attacks in 2001, said the non-profit organization Public Employees for Environmental Responsibility on 21 April. The statement follows a ruling last month by a judge at the US Labor Department that the EPA deliberately sought to conceal information about the case. The EPA reinstated Jenkins in May 2012, more than a year on from her dismissal, after an appeals board ruled that the EPA did not properly document its case (see *Nature* http://doi.org/34b; 2012).

New NASA deputy

After nearly two years, NASA has filled its vacant deputy-administrator post. Dava Newman was confirmed for the position by the US Senate on 27 April. As an aeronautics and astronautics scientist at the Massachusetts Institute of Technology in Cambridge, her research has focused on human space exploration, and she is credited with developing lightweight spacesuits to give astronauts more mobility. The position as NASA's second-in-command had been vacant since Lori Garver stepped down in September 2013.

↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Young film-goers are still viewing large amounts of smoking on screen, according to an analysis of 1,853 box-office hits released in the past 12 years. The report, by tobacco researcher Stanton Glantz at the University of California, San Francisco, and colleagues, shows that in 2014, 36% of films that can be viewed by children without an accompanying adult featured tobacco imagery. Although that was down from 68% in 2002, there was a greater frequency of tobacco incidents in films that did show smoking.

SMOKING ON FILM
The proportion of youth-rated blockbusters that contain tobacco imagery has fallen since 2002.
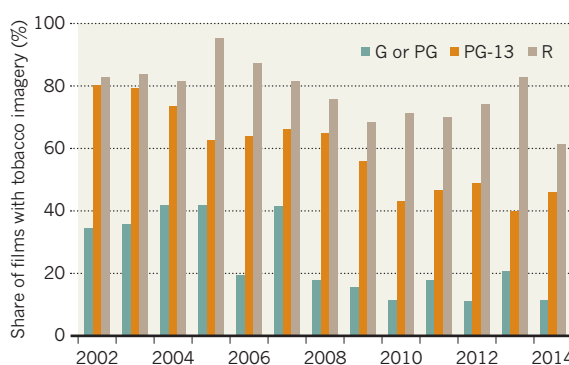
Ratings according to the Motion Picture Association of America. G, general audience; PG, parental guidance; PG-13, parents strongly cautioned; R, restricted.

NEWS IN FOCUS

YORGOS NIKAS/SPL
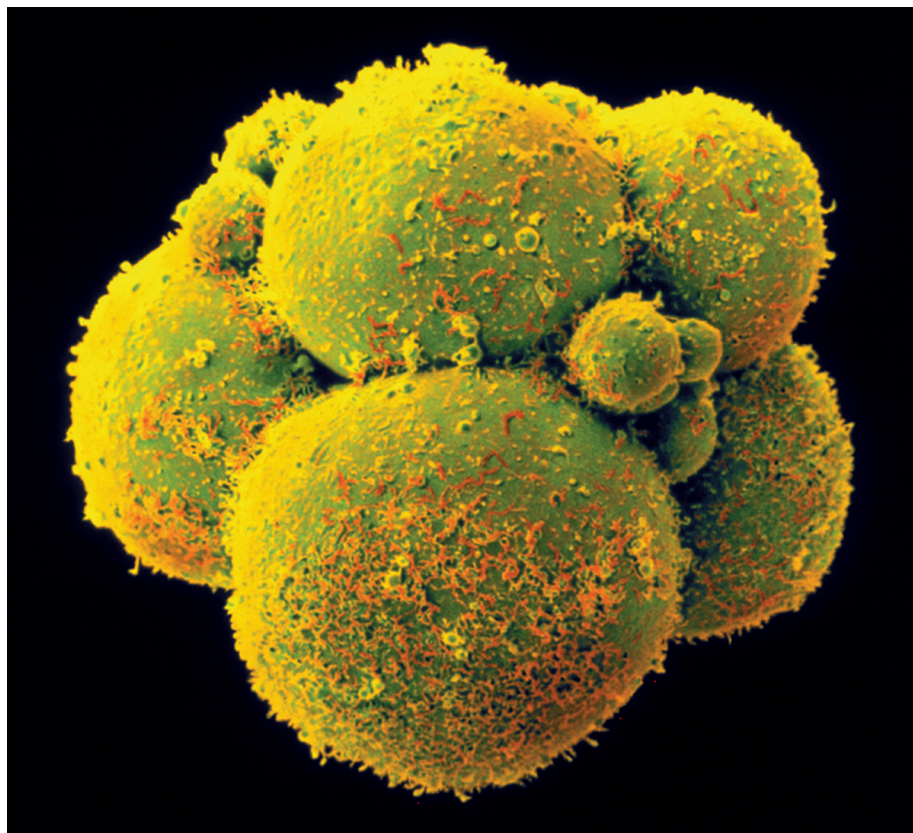
Human embryos are at the centre of a debate over the ethics of gene editing.

GENETIC MODIFICATION

Embryo editing sparks epic debate

In wake of paper describing genetic modification of human embryos, scientists disagree about ethics.

BY DAVID CYRANOSKI AND SARA REARDON

In a world first, Chinese scientists have reported that they have used powerful gene-editing techniques to modify human embryos. Their paper[1], published in the Beijing-based journal *Protein & Cell* on 18 April, came as no surprise to the scientific community, but it has ignited a wide-ranging debate about what types of gene-editing research are ethical. The publication also raises questions about the appropriate way to publish such work (see 'Publishing policy').

In the paper, researchers led by Junjiu Huang, a gene-function researcher at Sun Yat-sen University in Guangzhou, describe how they used a system of molecules called CRISPR/Cas9, known for its ease of use, to cut DNA in human embryos and then attempted to repair it by introducing new DNA.

In a deliberate attempt to head off ethical concerns, the team used non-viable embryos obtained from fertility clinics, in which eggs had been fertilized by two sperm and so could not result in a live birth.

Gene-editing techniques such as those that rely on CRISPR/Cas9 had previously been used to modify DNA in adult human cells and animal embryos. Earlier this year, rumours were circulating that the methods were being applied in human embryos too, but the Huang paper is the first published report of this.

The team used CRISPR/Cas9 to modify a gene that can cause a potentially fatal blood disorder called β-thalassaemia when it is mutated. Some researchers have suggested that such a procedure, if done in a viable embryo, could eradicate devastating genetic diseases before a baby is born. Others say that such work crosses an ethical line: in response to the rumours that the work was being carried out, researchers warned in *Nature*[2] and *Science*[3] in March that because the genetic changes to embryos — a procedure known as germline modification — are heritable, they could have an unpredictable effect on future generations.

Researchers have also expressed concerns that any gene-editing research in human embryos could be a slippery slope towards unsafe, unethical or non-medical uses of the technique.

SERIOUS OBSTACLES

Huang's team says that its results reveal serious obstacles to using the method in a clinical setting. The team injected 86 embryos with CRISPR/Cas9, along with other molecules designed to add in new DNA. The researchers then waited 48 hours, by which time the embryos would have grown to about eight cells each. Of the 71 embryos that survived, 54 were genetically tested. This revealed that just 28 were successfully spliced, and that only 4 of those contained the genetic material designed to repair the cuts. "That's why we stopped," says Huang. "We still think it's too immature."

His team also found a surprising number of 'off-target' mutations assumed to be introduced by the CRISPR/Cas9 complex acting on other parts of the genome. The effect is one of the main safety concerns surrounding ▶

PUBLISHING POLICY
Gene editing poses challenges for journals

Amid the discussion generated by a paper that reports gene editing in human embryos, the process behind its publication has also aroused curiosity.

Lead author Junjiu Huang of Sun Yat-sen University in Guangzhou, China, says that the paper, published on 18 April in the Beijing-based online journal *Protein & Cell*, was rejected by *Nature* and *Science*, in part because of ethical objections. Both journals keep details of their review processes confidential (*Nature*'s news team is editorially independent of its research editorial team), but acknowledge that gene-editing of human embryos is a complicated issue for them.

"This is a rapidly evolving and complex area for which we cannot — and should not — easily offer simplistic policies," says Ritu Dhand, editorial director for *Nature*. Nature Publishing Group is consulting with a range of experts to develop a "progressive policy" on the issue, she says.

Science, meanwhile, told *Nature*'s news team: "We believe strongly that the potential of genome editing must be viewed in terms of social mores and that the path forward must be developed through a consensus-building process."

The editors of *Protein & Cell* say that they published the paper to "sound an alarm" about such work. "In this unusual situation, the editorial decision to publish this study should not be viewed as an endorsement of this practice nor an encouragement of similar attempts," wrote Xiaoxue Zhang, managing editor at *Protein & Cell*, in an editorial published on 28 April (X. Zhang *Protein Cell* http://doi.org/35n; 2015). "We had serious discussion about the ethics of this issue," adds the journal's editor-in-chief, Zihe Rao. "We expected there might be difference of opinions, but it needs to be published to start discussion."

Springer, the publisher of *Protein & Cell*, confirmed that the journal had checked the researchers' institutional approval and the consent forms from the embryo donors. They also confirmed that the study was compliant with the Helsinki declaration on human-medical-research ethics and with Chinese law.

The paper sped through *Protein & Cell*'s review process: it was submitted on 30 March and accepted on 1 April. A spokesperson for Springer said that the paper was submitted with peer-review comments from *Nature* and *Science* and that the authors had made revisions on the basis of these, which facilitated the fast review. Another round of peer review was conducted in the two-day gap between submission and acceptance, said the spokesperson.

Two days is "quite long", says Rao. "You can e-mail the article to everyone at once. It's not like the old days." **Daniel Cressey and David Cyranoski**

that non-viable embryos are discarded," says John Harris, a bioethicist at the University of Manchester, UK. "I don't see any justification for a moratorium on research," he adds. Church, meanwhile, notes that many of the earliest experiments with CRISPR/Cas9 were developed in human induced pluripotent stem cells, adult cells that have been reprogrammed to have the ability to turn into any cell type, including sperm and eggs. He questions whether Huang's experiments are any more intrinsically problematic.

Modifying human embryos is legal in China and in many US states. Asked whether Huang's study would have been funded under its rules, the US National Institutes of Health says that it "would likely conclude it could not fund such research", and is watching the technology to see whether its rules need to be modified.

Because the embryos Huang's team used were initially created for *in vitro* fertilization, not for research, the work would already have overcome many of the ethical hurdles it would face in other countries too, adds Tetsuya Ishii, who studies bioethics and policy at the University of Hokkaido in Sapporo, Japan.

NEXT STEPS
Applying gene editing to human embryos could answer plenty of basic scientific questions that have nothing to do with clinical applications, says George Daley, a stem-cell biologist at Harvard Medical School, who supports editing of human embryos *in vitro* for research purposes.

For instance, altering developmental genes with CRISPR/Cas9 could help to reveal their functions. "Some questions about early human development can only be addressed by studying human embryos," he says.

Gene editing could also be used to engineer specific disease-related mutations in an embryo, which could then be used to produce embryonic stem cells that could act as models for testing drugs and other interventions for disease, says Daley.

Huang now plans to work out how to decrease the number of off-target mutations using adult human cells or animal models.

Still, researchers expect to see more gene-editing studies in human embryos. "The ubiquitous access to and simplicity of creating CRISPRs," says Lanphier, whose company applies gene-editing techniques to adult human cells, "creates opportunities for scientists in any part of the world to do any kind of experiments they want." He expects that more scientists will now start work on improving the results of the Huang paper. A Chinese source familiar with developments in the field said that at least four groups in China are pursuing gene editing in human embryos. ∎

▶ germline editing because these unintended mutations could be harmful.

The rates of such mutations were much higher than those observed in gene-editing studies of mouse embryos or human adult cells. And Huang notes that his team probably detected just a subset of the unintended mutations because their study looked at only a portion of the genome known as the exome. "If we did the whole genome sequence, we would get many more," he says.

Huang wonders whether there might be something intrinsically different that makes the human embryo more susceptible to extra mutations than animal embryos are. Another possibility — suggested by some critics of the work, he says — is that CRISPR/Cas9 worked differently in the embryos that his team used because they were the product of two sperm fertilizing an egg.

For some, these technical challenges support arguments for a moratorium on all research on human germline modification. "I think the paper itself actually provides all of the data that we kind of pointed to," says Edward Lanphier, president of Sangamo BioSciences in Richmond, California, and a member of the group that wrote the *Nature* article[2] calling for a moratorium.

But George Church, a geneticist at Harvard Medical School in Boston, Massachusetts, disagrees that the technology is so immature. He says that the researchers did not use the most up-to-date CRISPR/Cas9 methods and that many of the researchers' problems could have been avoided or lessened if they had.

"Some questions about early human development can only be addressed by studying human embryos."

Although researchers agree that a moratorium on clinical applications is needed while the ethical and safety concerns of human-embryo editing are worked out, many see no problem with the type of research that Huang's team did, in part because the embryos could not have led to a live birth. "It's no worse than what happens in IVF all the time, which is

1. Liang, P. *et al. Protein Cell* http://dx.doi.org/10.1007/s13238-015-0153-5 (2015).
2. Lanphier, E. *et al. Nature* **519,** 410–411 (2015).
3. Baltimore, D. *et al. Science* **348,** 36–38 (2015).

Canadian Prime Minister Stephen Harper (left) and finance minister Joe Oliver announced the budget.

POLITICS

Canada pushes applied research

Budget increases government partnerships with industry, but plays down basic science.

BY MARGARET MUNRO

With six months to go before the next Canadian election, the reigning Conservative government has introduced a budget that emphasizes applied research and scientific collaboration with industry.

The 518-page proposal, released on 21 April, will take effect in the coming weeks. It spells out how Prime Minister Stephen Harper's government plans to balance its budget this year — at the same time pledging plenty of new spending in years ahead.

The Canada Foundation for Innovation would receive Can$1.33 billion (US$1.09 billion) in new money for university and hospital research facilities, to be doled out over six years beginning in 2017. The budget also includes a modest 2% hike for the country's research-funding councils, much of it targeted for areas "that will fuel economic growth".

David Robinson, executive director of the Canadian Association of University Teachers in Ottawa, calls the plan a missed opportunity. "The government continues to miss the fact that real innovation and scientific advancements are driven by long-term basic research, not short-term market demands," he says.

> *"There's a trend to use universities as surrogates for industrial research."*

Much of the promised funding for research is destined for industrial partnerships. This is in keeping with other Conservative policies, some of which have strained relationships between the government and scientists.

The Harper government has been criticized for micromanaging how research money is spent and for how science is communicated. "Muzzled" federal scientists are no longer free to discuss their research findings with the media without clearance from the government (see *Nature* **483,** 6; 2012). And funding for basic-research and federal science programmes such as the Experimental Lakes Area, an acclaimed freshwater research centre in northwestern Ontario, has been cut as the Conservatives have boosted spending on applied research.

"There's a trend to use universities as surrogates for industrial research," says Paul Dufour, a fellow at the Institute for Science, Society and Policy at the University of Ottawa, and a science-policy consultant.

The trend continues in the latest budget. The proposal includes Can$119 million over the next two years for the National Research Council's "industry-partnered research and development activities"; Can$46 million per year, starting in 2016, to the research-granting councils in part to "support their growing interactions with the private sector"; and Can$56.4 million over four years, starting in 2016, for business-related internships for graduate students.

Although the budget allocates just Can$10 million in new money this year for the government's "world-class advanced research" initiatives — which include ▶

➡ **MORE ONLINE**

CHEMICAL WEAPONS

Nations reflect on a century of chemical warfare go.nature.com/lhpeqw

MORE NEWS

● Invasive lionfish discovered in Brazil go.nature.com/magbmp
● Cosmic rays reveal the secrets of thunderstorms go.nature.com/kjkllm
● NASA probe set to end mission with dramatic crash into Mercury go.nature.com/ayvdgv

PSYCHOLOGY

Experiment offers the illusion of being invisible go.nature.com/6x6oib

▶ projects in physics, space technology and astronomy — it pledges to invest more than Can$200 million per year in new and expanded programmes starting in 2016.

Dufour says that the research landscape has become "a bit of a dog's breakfast" with the Conservatives' targeted research initiatives. And he notes that the focus on industrial research has not improved Canada's overall research and development performance.

The Organisation for Economic Co-operation and Development (OECD) reported in November that Canada has dropped out of the top ten research and development performers, ceding ground to more aggressive countries such as Taiwan. The OECD also reports that Canada has seen one of the sharpest declines in the percentage of gross domestic product spent on research.

ELECTION TOPIC

Opposition parties hope to make science an issue in the October election, but their criticism might be blunted by initiatives such as Harper's Can$243-million, ten-year commitment to the Thirty Meter Telescope (TMT), announced earlier this month and highlighted in the budget. The international observatory is to be constructed on the summit of Mauna Kea in Hawaii.

"You can rest assured the TMT will be trotted out during the campaign as evidence that they are investing in research that is probing the origins of the Universe," says Scott Findlay, a biologist at the University of Ottawa and founding member of Evidence for Democracy, a group that pushes for more science-based decision-making in government.

The budget also commits Canada to supporting the International Space Station until 2024, an extension that the United States and Russia have already committed to.

As the election nears, the government is expected to start announcing projects financed by its Can$1.5-billion Canada First Research Excellence Fund, which aims to put Canadian universities "among the best in the world for talent and breakthrough discoveries". Applicants must spell out the strategic relevance of proposed projects and their potential to "create long-term economic advantages for Canada".

Many of Canada's university presidents enthusiastically endorsed the fund when Harper announced details in December. And some were quick to applaud the new budget's provisions. "They will benefit Canada and Canadians now and for years to come," says David Barnard, president of the University of Manitoba and chair of the Association of Universities and Colleges of Canada. ∎

Pope Francis says that humanity "has slapped nature in the face".

SOCIETY

Vatican hosts climate meeting

Religious leaders and scientists gather to discuss moral implications of global warming as Pope drafts key letter.

BY EDWIN CARTLIDGE

Pope Francis has drawn attention for his progressive teachings on sexuality, contraception and abortion. Now he and the Catholic Church are tackling another contentious issue: climate change.

This week, scientists, religious figures and policymakers gathered at the Vatican to discuss the science of global warming and the danger posed to the world's poorest people. The meeting came as Francis prepares an encyclical letter to bishops on climate change for release this summer, ahead of United Nations climate negotiations in December. The Pope's strong feelings on the matter are apparent: in January he said that people were "mostly" responsible for recent warming and that they have "slapped nature in the face".

Two scholarly Vatican groups that advise the Pope on scientific issues organized the meeting. One, the Pontifical Academy of Sciences, has held several related conferences, including a 2011 meeting on melting glaciers and another last year on sustainable development. But the climate gathering this week is the first at the Vatican to include religious leaders as well as scientists.

Roughly 20–25 individuals representing religions such as Protestant and Orthodox Christianity, Judaism and Islam joined a similar number of scientists. Nobel Prize-winning chemist Paul Crutzen, climatologist James Hansen and former UK Royal Society president Martin Rees were scheduled to attend, along with UN secretary-general Ban Ki-moon and Italian President Sergio Mattarella. All participants were asked to sign a statement "on the moral and religious imperative of sustainable development".

Peter Raven, a botanist at the Missouri Botanical Garden in St Louis and an organizer of the meeting, says that such messages are crucial for winning public support for policies to reduce greenhouse-gas emissions. "The only way politicians can put binding agreements in place is if they are convinced that enough people care about the problem," he

says. "And that will only happen if the problem has a strong ethical component."

Among the moral arguments for action on climate change is the idea that society must preserve Earth for future generations, says another meeting organizer, climate scientist Veerabhadran Ramanathan of the Scripps Institution of Oceanography in La Jolla, California. A further argument is that the world's 3 billion poorest people produce a fraction of humanity's annual greenhouse-gas emissions — about 5% — but are predicted to suffer disproportionately from the impacts of climate change, such as more extreme weather.

Francis is not the first major religious leader — nor even the first pope — to speak out on climate change. In a 2009 statement, Archbishop of Canterbury Rowan Williams and other UK religious representatives recognized a moral imperative to address the causes of global warming. And last month, the presiding bishop of the US Episcopal Church, Katharine Jefferts Schori, described those who attribute global warming to purely natural causes as "often driven by greed and self-centred political interests, and sometimes by wilful blindness".

Closer to home, Francis's predecessor Benedict XVI made numerous statements on the environment, describing climate change in 2011 as a "worrying and complex phenomenon". But Ramanathan thinks that the huge popularity of the current pope presents a unique opportunity to raise awareness. "Pope Francis has become a Kennedy-like figure who goes beyond Catholicism and appeals to the whole world," he says.

Even a popular pope can stir controversy, however. Francis's support for action on climate change has raised hackles among some conservative Catholics, such as Steve Moore, who is chief economist at the Heritage Foundation, a free-market think tank in Washington DC. Francis has "allied himself with the far left", Moore says.

But Dan Misleh, executive director of the Catholic Climate Covenant, an environmental group in Washington DC, says that Francis is unlikely to endorse specific policies to fight climate change. "I don't think he is going to say that a carbon tax is preferable to a cap-and-trade programme," Misleh says. "But I think he will say that there is a connection between how we treat the planet and how we treat one another."

Such a strategy will allow Francis to avoid seeming "ideological" while still speaking

"I think he will say that there is a connection between how we treat the planet and how we treat one another."

clearly about the role of humans in climate change, says James Bretzke, a Catholic theologian at Boston College in Massachusetts. Already, he says, several Vatican figures who are likely to be involved in drafting the Pope's encyclical have made unequivocal statements on global warming. Last year, for example, Bishop Marcelo Sánchez Sorondo, chancellor of the Pontifical Academy of Sciences, said that "essentially all documents published now accept as a scientific truth that climate change is due to human activity".

The document set to be signed at this week's meeting contains a very similar statement, according to co-organizer Partha Dasgupta, an economist at the University of Cambridge, UK. Attendees were scheduled to discuss the physical and chemical processes that underlie global warming, he says, but a debate on the idea that human activities drive that warming was not planned. "Religious leaders might ask whether other scientists have different views," says Dasgupta, adding that the small minority of scientists who are sceptical of climate change were not represented at the meeting.

Still, the Pope will be careful to point out where scientific uncertainties remain, Bretzke predicts. Such caution is borne of the Vatican's famous over-certainty in the past. "Centuries ago, the Church got on the wrong end of cosmological arguments," he says. "It has been chastened by that." ∎

Gene–testing firms set sights on drug development

Companies race to gather large DNA data sets in bid to find treatments based on genetics.

BY ERIKA CHECK HAYDEN

Companies that offer genetic testing directly to consumers are renewing their ambitions. Recent moves by US regulators have given the firms fresh hope that the large genetic data sets they amass will have commercial as well as scientific value, spawning diagnostic tests or drugs.

The moment seems ripe. In February, the US Food and Drug Administration (FDA) allowed 23andMe — a company in Mountain View, California, that has offered genome analysis to consumers for nearly a decade — to begin marketing a test for mutations that cause a rare disease called Bloom syndrome. This was the first time that the agency had approved a genetic test marketed to the public, not clinicians — and companies are betting that it will not be the last.

Many expect the FDA to further expand the types of medically relevant information that consumer genetic tests can supply. The February decision "is a huge step for the field", says Ken Chahine, a senior vice-president at Ancestry.com DNA of Provo, Utah, which offers genetic testing to the public. "You're starting to see others get into this space because everyone sees a lot of value in it."

Ancestry.com DNA, a subsidiary of genealogy website Ancestry.com, is one of those newcomers. The company has so far collected DNA from 850,000 customers, which it uses to help people to find relatives among members of its genealogy-focused social network. It is now exploring the idea of collecting users' medical histories, too, so that it could offer them medically relevant information and investigate the genetic roots of diseases.

23andMe is farther down this road and has already attracted interest from health-care firms. In January, it signed a US$60-million deal with California biotechnology company Genentech, which will use genetic data from 23andMe customers to develop therapeutics. The personal-genomics firm has also hired former Genentech executive Richard Scheller to lead an in-house effort to develop drugs. And, similar to Ancestry.com, 23andMe has access to a home-grown social network that encourages its customers to interact with the company and each other.

Commercial firms are not the only group seeking to build large pools of participants for genetic studies — or the only ones harnessing social networks for this cause. The medical-research study Genes for Good, which launched on 31 March, is seeking ▸

▶ participants through Facebook. Users of the social-networking service can download a Genes for Good app that guides them through a series of health surveys to collect basic information, such as height, weight and whether a person uses tobacco. Users who complete 15 of these surveys can receive a 'spit kit' to collect saliva for DNA sequencing, and are encouraged to continue providing health information afterwards.

"We wanted to do something that would give us the possibility of capturing very large numbers of people in a way that is more engaging than the traditional types of study," says Gonçalo Abecasis, a statistical geneticist at the University of Michigan in Ann Arbor who is leading the study.

WIDENING THE NET

Drawing on 23andMe's example, the project will return genetic information to users — such as hints about ancestry drawn from a user's genetic markers. The project also hopes to share the information it collects with other academic scientists,

"They're smart to tether their research to a platform that has a much broader market base of people."

or even companies developing new therapies, after removing identifying information such as names and addresses.

So far, Genes for Good has recruited 4,200 potential participants of the 20,000 it seeks, and hundreds of those have already sent back their spit kits. That is a relatively small number of genomes compared to the amount of genetic information contained in private databases. And it is dwarfed by the amount of data being amassed by public projects, such as one run by the UK Department of Health that aims to sequence the genomes of 100,000 patients.

Such efforts, in theory, could compete for participants with commercial ventures such as Ancestry.com DNA and 23andMe, which ask customers to pay for testing similar to the type that Genes for Good will offer for free. It is not clear whether that will affect the companies' ability to attract customers and expand their data sets, given their significant head start.

Genes for Good is tapping the power of Facebook, with its 1.44 billion active monthly users. Given this potential, the project could become more massive than any previous genetic study, says Michelle Meyer, a bioethicist and legal scholar at the Icahn School of Medicine at Mount Sinai in New York. "They're smart to tether their research to a platform that has a much broader market base of people than we're used to seeing in research studies," she adds. ∎

DNA from humans who lived in the Andes 9,000 years ago gives clues to how South America was peopled.

ARCHAEOLOGY

South America settled in one go

Tests of human remains from a cave high in the Peruvian Andes point to a single early migration.

BY EWEN CALLAWAY

Humans settled in South America in a single wave of migration not long after their ancestors first crossed from Siberia into the Western Hemisphere during the last ice age, genetic evidence suggests. The finding, based on DNA from the remains of five ancient humans who lived high in the Peruvian Andes, also hints at how ancient Andeans evolved to thrive at altitudes of more than 4,000 metres.

Presented on 16 April at the annual meeting of the Society for American Archaeology (SAA) in San Francisco, California, the research sheds light on the last major continental migration in human prehistory — and one of the least understood.

South America's oldest-known human occupation site — the 14,600-year-old Monte Verde settlement in Chile — suggests that people quickly reached the continent after crossing the Bering land bridge one or two thousand years earlier, perhaps skirting the Pacific coast. But some argue for a second migration. They use skeletal evidence to propose that the long, narrow skulls of South Americans living more than about 5,000 years ago differ too much from the rounder heads of more-recent

inhabitants, and of living indigenous people, to represent one continuous population[1].

In the 1950s and 1960s, Peruvian archaeologist Augusto Cardich discovered human remains bearing the characteristic ancient skull shape in a rock shelter perched high in the Andes in a region called Lauricocha. Carbon dating put the site at around 9,000 years old, making it a classic archaeological example of early high-altitude living. Later researchers dated remains from Lauricocha closer to 5,000 years old, greatly diminishing its appeal for researchers interested in the early prehistory of South America. "People simply forgot about this site. Nobody was interested any more," says Lars Fehren-Schmitz, a biological anthropologist at the University of California, Santa Cruz.

More than half a century after the site's discovery, Fehren-Schmitz's team got permission to look at five human skeletons excavated from Lauricocha and held at the National Museum of the Archaeology, Anthropology and History of Peru in Lima. The team redated the remains, remeasured the skulls and extracted DNA.

Their work, presented by Fehren-Schmitz at the SAA meeting, paints a complicated picture of Lauricocha. Two of its residents, a woman and a 2-year-old child, died nearly 9,000 years ago. The third, a man, perished around 2,500 years later and another man died

about 2,300 years later still. The fifth specimen was not dated because of its condition. Only the woman's skull had a long, narrow shape, which is known as dolichocephaly.

To gauge whether the Lauricocha residents were descended from members of more than one migration, the team sequenced the DNA from their mitochondria, cellular organelles that are passed directly from mother to child, and thus trace maternal ancestry. All five people were found to be descended from maternal lines that are common among modern and ancient indigenous people of North and South America. The males' Y chromosomes put them on a lineage that arose in the region around the Bering Strait some 17,000 years ago — the most widely accepted time and place for the original human migration into the Americas. These and other DNA data suggested that all the Lauricocha humans are descended from the first humans to reach the Americas, supporting one migration into South America.

Not everyone accepts the argument. Such a wide-ranging conclusion cannot be drawn from a handful of remains found at a single site, says Tom Dillehay, an archaeologist at Vanderbilt University in Nashville, Tennessee. He argues that the 5,000-year gap between the earliest inhabitants at Monte Verde, which Dillehay excavated, and the earliest Lauricocha skeletons leaves the door open for earlier migrations.

"There may well be a single migration — they don't have the evidence for it," he says.

Ancient human DNA from other parts of South America, such as the Amazon, may yet point to a second migration into the continent, agrees Fehren-Schmitz. His team is now analysing full genomes from Lauricocha and other South American sites to obtain a fuller picture of the continent's past.

Deborah Bolnick, a biological anthropologist at the University of Texas at Austin, accepts the argument for a single migration. But regardless of the number of waves, there would have been further population movements after that initial, southbound founding trek, she says, adding that analysis of full genomes from ancient South Americans will give a more nuanced picture of the continent's prehistory. Researchers studying ancient DNA have found evidence for a massive migration from present-day Russia and Ukraine into western Europe around 4,500 years ago[2], and Bolnick says that similar upheavals are bound to have happened in the Americas. In a 2014 paper, for instance, Fehren-Schmitz documented a migration into the Central Andes about 1,400 years ago, possibly driven by drought in lower-lying areas[3].

Ancient South American genomes may also show how humans adapted to the New World. In a separate study, Fehren-Schmitz and his colleagues looked at a gene variant that protects against altitude sickness. Between around 8,500 and 600 years ago, the presence of the variant increased markedly among the Andeans.

Fehren-Schmitz says that the finding is merely suggestive of local adaptation. But with full, ancient genomes in hand, researchers will be able to look more thoroughly for signs of adaptation to high altitude, says Bolnick. With that question in mind, her team is sequencing DNA from remains of mountain residents of Argentina. Ancient genomes from the Americas may also reveal how humans adapted to shifts in diet — early Americans domesticated maize (corn), potato and other crops — and to diseases imported from Europe, such as smallpox. South American prehistory, it seems, is about to get a lot more complicated. ∎

1. Walter, A., Neves, W. A. & Hubbe, M. *Proc. Natl Acad. Sci. USA* **102,** 18309–18314 (2005).
2. Callaway, E. *Nature* **518,** 284–285 (2015).
3. Fehren-Schmitz, L. *et al. Proc. Natl Acad. Sci. USA* **111,** 9443–9448 (2014).

CORRECTION
The Editorial 'Highway to health' (*Nature* **520,** 407; 2015) stated that the African Centres for Disease Control and Prevention would need at least $300 million and 60 staff members. This should have read $60 million and 300 staff.

COLLATERAL DAMAGE

How a case of misconduct brought a leading Japanese biology institute to its knees.

BY DAVID CYRANOSKI

Hironobu Fujiwara was already troubled as he made his way to work on 5 August 2014. As a laboratory head at the RIKEN Center for Developmental Biology (CDB) in Kobe, Japan, Fujiwara had been enduring criticism ever since a case of misconduct had been exposed at the centre earlier that year. A media furore had escalated week by week, as newspapers, social media and television programmes all demanded an explanation for how scientists at a prestigious institute could have conducted shoddy work. Although he was not involved in the suspect studies, Fujiwara and many other employees at the centre felt under attack.

And then, when he arrived that day, he heard terrible news. Yoshiki Sasai — a founding member of the CDB who had been implicated in the misconduct case — had hanged himself in an adjoining building that morning. "I was just in shock," Fujiwara says. "I didn't know what to think, or whether to believe it was true."

Revelations of scientific misconduct always cause collateral damage: they taint colleagues and co-authors of the person responsible, and can close down labs. But the case at the CDB triggered unusually strong and far-reaching aftershocks.

The case involved two high-profile papers in *Nature*[1,2] that described a surprisingly simple method to make pluripotent stem cells — cells with the prized ability to develop into any of the body's cell types. After the misconduct came to light, a crucial report blamed not just the individuals involved, but also the entire centre — and recommended that it be dismantled. Since then, the CDB's funding has been slashed, half of its labs have been closed, merged or assigned to other organizations, and its leadership has been replaced. The upheaval reaches far beyond the centre. A government science administrative reform has been put on hold and the scientific community across Japan is now bracing for the impact of anti-misconduct policies introduced in the wake of the affair.

To many scientists and journalists in Japan, this was an appropriate response. They thought that the stem-cell results had been sensationalized, and that it was fitting to take drastic action when those results were shown to be false.

But other scientists in Japan and many abroad call it an overreaction and say that events over the past year reveal how responses to a crisis can create their own problems. The way that scientists and the media reacted exposed long-standing jealousies in Japan towards the 15 well-funded RIKEN centres and institutes, which had already made the CDB a target for critics. The frenetic media and social-media response created a storm that bureaucrats and science-policy administrators were desperate to quell. And reporting by *Nature* calls into question whether the CDB was given enough of a chance to defend itself: a committee behind the key report on the future of the CDB did not interview most of those involved to find the causes of the misconduct. Teruo Kishi, who led the committee, defends the process. "You seem to think that by talking to people involved we would find something out," he says. "But we wouldn't have learned anything from asking them."

Some scientists worry that the ferocity of the response will hamper research for current or former CDB scientists and even dampen Japan's ability to support innovative science in the long run. "It was only two papers, but from the beginning to the end the media kept blowing it up larger and larger," says former cancer geneticist Yuko Ito, now a science-policy expert at the Japan Science and Technology Agency in Tokyo. "In the end, a lot of scientists became victims."

SCIENTIST'S PARADISE

RIKEN was founded in 1917 as the Institute of Physical and Chemical Research in Tokyo, and it expanded heavily into biological sciences with a series of institutes that opened from the late 1990s onwards. The centres became known in Japan as a scientist's paradise, because researchers there had no undergraduate teaching responsibilities and enjoyed generous salaries and research funds that meant that they did not depend on grants. Even before the stem-cell crisis, "the relationship between RIKEN and universities was not really good", says Hiroshi Hamada, a developmental biologist formerly at Osaka University, who took over as CDB director this year.

Even among RIKEN centres, the CDB stood out. Launched in 2000, when Japan was trying to rejuvenate its research infrastructure, the institute abandoned the crusty hierarchical structure that encumbered university laboratories. Scientists and staff address each other with the common *san* rather than the *sensei* typically used for superiors. In a 2002 interview with *Nature*, Sasai vowed to "give young researchers a degree of independence" that was previously unknown in Japan (see *Nature* **415,** 952–953; 2002) — and soon young principal investigators, some still in their 20s, were mapping out their research programmes. A handful of established researchers, known as group directors and including Sasai and founding director Masatoshi Takeichi, oversaw operations.

The approach quickly paid off, and high-profile results emerged. Lab leader Mitinori Saitou won acclaim for engineering germ cells in a dish (see *Nature* **500,** 392–394; 2013); Sasai built a reputation for his skill growing eye and brain structures in culture (see *Nature* **488,** 444–446; 2012). In a country where research

has often struggled to make an international impact, the CDB developed a global reputation for exceptional work in developmental biology. The institute was proud to state that of 163 papers that its researchers published in 2013, one-third were in leading international journals including *Nature*, *Science* and *Cell*.

The centre seemed to score another big coup in January 2014, when some of its scientists published the two papers on pluripotent stem cells. Developing the technology, called stimulus-triggered acquisition of pluripotency, or STAP, was the project of biochemist Haruko Obokata, who had started the work at Harvard University in Cambridge, Massachusetts, and brought it to the CDB. She had worked with three highly respected scientists at the centre: mouse-cloning pioneer Teruhiko Wakayama, stem-cell biologist Hitoshi Niwa and Sasai.

When the technique was published, it caused excitement in Japan. Pluripotency had become almost a household word since Shinya Yamanaka at Kyoto University had won a Nobel prize in 2012 for his work on a related technique to make induced pluripotent stem (iPS) cells. And the media made much of Obokata, a young, quirky woman who wore a *kappogi* — a type of traditional apron that she had received from her grandmother — instead of a lab coat, and whose laboratory walls were painted pink and yellow. She was everything that the traditional Japanese scientist was not.

Within weeks of the papers' publication, however, the work began to unravel. Science blogs pointed to manipulated figures and other scientists could not repeat the results (see go.nature.com/h9tr5w). A committee at RIKEN investigated and found signs of

misconduct. RIKEN officials held marathon four-to-five-hour press conferences, formally accused Obokata of misconduct on 1 April, and recommended that the papers be retracted, which they later were. The committee's report said that Sasai and Wakayama had not been involved in the misconduct, but shared "grave responsibility" for not catching the problematic data. Niwa was cleared of wrongdoing.

At this point, the typical university fraud case in Japan — and there have been several high-

"IN THE END, A LOT OF SCIENTISTS BECAME VICTIMS."

profile examples over the past 15 years — winds down. The researcher implicated in the misconduct usually resigns and the tainted papers are retracted. Witness, for example, the resignation of University of Tokyo molecular biologist Shigeaki Kato and the retraction of dozens of his papers between 2012 and 2014. In that case, the damage was limited to members of Kato's group.

But with STAP, the media was hungry for more. News teams camped in the lobby of the

CDB and filled the corridors of the science ministry, looking for twists to the story. They trailed Obokata, Sasai and many other CDB researchers and administrators. (Japan's national broadcaster, NHK, apologized after Obokata was injured in one scrape with reporters.)

News articles, tweets and blogs started spreading the blame to the CDB and RIKEN as a whole. Critics said that the misconduct investigation had been hastily prepared, and had failed to get to the bottom of the problem. They also focused on a patent that RIKEN and the scientists involved in the work had filed on the STAP technique — standard practice in academia, but taken as evidence that the centre and its researchers were driven by economic incentives. Noriko Osumi, a developmental biologist at Tohoku University and one of the most vocal critics of RIKEN during the STAP crisis, faulted the organization for "inviting excessive attention from the media", which she blamed on "the influence of commercialization and industrialization of research in the life sciences". The desire for money and media attention was seen by others as a reason that senior scientists at the CDB had thrown their weight behind the STAP project and let their guard down, allowing poor science to take place.

Young CDB investigators who had nothing to do with Obokata's work were worried that their reputations were being tarnished. "At first the line was blurred, then we saw that there was no line at all," says Yu-Chiun Wang, who had arrived at the CDB in October 2013 from Princeton University in New Jersey. "We were all seen as part of the same criminal complex." CDB scientists, technical staff and secretaries were being slandered online, says Takeshi Imai, a neurobiologist at CDB. "Myself, my lab members and colleagues were also there," he says. Even researchers at other RIKEN centres felt under attack.

QUIETING THE STORM

The media hullaballoo was a headache for the science ministry, which was making plans to give certain research organizations unprecedented autonomy. RIKEN was to be the first. The ministry "wanted to finalize the misconduct matter quickly", says Maki Kawai, a former RIKEN executive director. It thought an aggressive action "would seem a more reasonable response to the taxpayers' voice, which is the media".

On 9 April 2014, shortly after Obokata was charged with misconduct, RIKEN established an independent 'reform committee' chaired by Kishi, a lean 75-year-old materials scientist with a long list of top administrative positions. On 12 June, the Kishi committee released eight recommendations, including the promotion of research integrity, new fraud-prevention measures and a more thorough investigation of the STAP papers. But among these, one recommendation leapt out: to "dismantle" the CDB. "The Kishi report threw me out of my chair,"

Teruo Kishi (second left) of the RIKEN reform committee, at a 2014 press conference on the CDB's future.

says Ichiro Hiratani, a developmental biologist who had joined the CDB the previous year.

The Kishi committee's main task was to suggest measures to prevent misconduct, but half of its report analysed what had happened in the STAP misconduct episode, and why. The centre should be dismantled, the report said, because "the CDB, as an organization, had structural flaws, that induced or could not deter fraud". One of the main problems was an ossification of leadership: since its founding, the centre had had basically the same group directors. These, the report said, had come unconsciously to rely on each other in a cosy relationship that prevented critical scrutiny of each other's decisions.

Takeichi says that he and the CDB had been aware that a change was overdue well before the crisis. Over the preceding three years, the centre had searched for a foreign director to breathe new life into the CDB and increase its global status, but found it difficult to recruit a scientist of suitable calibre. However, Takeichi's leadership had not been in question before the STAP crisis. In 2011, RIKEN's advisory committee opposed his retirement.

The Kishi report said that poor governance at the CDB had led to mistakes. It accused the centre of circumventing normal procedures to help hire Obokata, and "surmises" a reason: "the CDB was strongly motivated by a desire to acquire a groundbreaking result" that would outdo Yamanaka's discovery of iPS cells. The report "surmises" that Sasai "naturally was involved in STAP in anticipation of the huge budget" that it would bring to the CDB. The report also said that Obokata's decision to wear a *kappogi* was part of a "showy PR strategy" orchestrated by Sasai.

In response to the report, RIKEN quickly convened a committee to start working out how to implement the recommendations. "Given how harshly the media was treating RIKEN, it had no choice but just accept it," a science-ministry official who does not want to be named told *Nature*. But Takeichi and Sasai contested many of the assertions in the report. Takeichi said that the hiring process for Obokata was normal; he and Sasai denied that there was a desire to outdo the iPS discovery, and other researchers at the CDB say that there was no pressure to do so. Part of Sasai's job as CDB deputy director was to organize and seek funding — a job at which he was very successful — but in an e-mail to *Nature*'s news team in June 2014, he denied that he had become involved in STAP because it would attract money to the institute. Sasai, Takeichi and RIKEN's public-relations office have all denied being involved in Obokata's choice of clothes.

In the interview with *Nature* in February, Kishi stood by his committee's report and acknowledged that it went beyond fact. "There was a lot of speculation, guesses, in the report," he said. "But they were guesses made with confidence," on the basis of what he had seen in long, televised press conferences.

Kishi said that in preparing its report, the committee did not talk to Obokata, Sasai or anyone at the CDB aside from Takeichi — a fact that other members of the committee confirmed. Kishi says that RIKEN would not let the committee meet with Sasai. But, Kishi said, "his televised conference made me feel that I didn't need to hear any more than what he said. I felt he said enough, he cannot be changed."

"GETTING ALMOST NO SUPPORT FROM SCIENTISTS WAS SHOCKING AND DEPRESSING."

Then Kishi passed a surprisingly harsh judgement on Sasai. "Even if we had asked him, he would not have told us the truth," he said. This differs from the opinion of those who knew Sasai well, who say he was an honest, broad-minded person, devoted to scientific research.

Committee member Masaki Nakamura, a research-integrity specialist and historian of science at Osaka University, also defended the report. He says that speculation was used in it the same way that a prosecutor will fill in the motivations of a defendant in a court of law. Four other members of the Kishi committee did not respond to requests for comment, or refused to comment. Nicholas Steneck, an expert in research integrity at the University of Michigan in Ann Arbor, says that in a misconduct case, he personally "would not be comfortable speculating about anything that was not based on some facts and direct information". He says, however, that such speculation is surprisingly common among scientists. "History and reporting on current events should follow basically the same methods as science," he says.

When the Kishi report came out, many foreign scientists found its conclusions excessive and arbitrary, and more than 150 letters of support were sent to the CDB. But in Japan, it was mostly received uncritically. Scientific organizations such as the Science Council of Japan threw their weight behind the judgement. The most dispiriting thing for many CDB researchers was the indifference or criticism from their scientific colleagues — some of it a reflection of the resentment of the CDB and RIKEN that had built up over the years. "Newspaper and

media, fine, they are trying to sell newspapers," says Hiratani. "But getting almost no support from scientists was shocking and depressing."

The fallout was especially hard on Sasai. Until June he had been "somehow getting by", says Keiko Muguruma, a stem-cell biologist at the CDB and a frequent collaborator with Sasai — they had just submitted a paper for publication. "In science, he felt he could recover from the harm to his reputation. But for things that he had no control, like the dismantling and budget decrease, which would affect all the young researchers, he felt guilt and responsibility," says Muguruma. According to the lawyer working with Sasai's family, the Kishi report and media attacks were factors in explaining Sasai's suicide. Kishi and Nakamura respond that it is difficult to know why Sasai took his own life.

In August, the CDB began making changes in line with the Kishi report. RIKEN announced a plan to introduce new fraud-prevention measures and strengthen governance.

Months later, in November, Takeichi stepped down and an interim director was put in place. The CDB saw 9 of its 40 laboratories shifted to other RIKEN centres, and another 11 were merged or closed. The centre also changed its Japanese name to "centre for research into the formation of multicellular systems", although it kept its original English name. In December, after announcing that she had failed to repeat her experiments, Obokata resigned.

The changes are still going on. On 1 April this year, Hamada took over as director and a 40% cut to the centre's budget kicked in. Researchers have been scrambling to supplement their finances with grants.

ASSESSING IMPACT

In interviews in the past two months, Kishi and Nakamura expressed tempered views of the CDB. Nakamura says that Takeichi's leadership had been "extremely good in comparison to other universities and research institutes". Although "we wrote about problems in their research ethics and education", Nakamura says, the CDB "was relatively speaking really quite advanced".

The word dismantle (*kaitai*) was a strategic choice meant to please an angry press rather than to put an end to the CDB, say both Kishi and Nakamura. It was "an appeal to society, trying to show the idea that RIKEN was taking this problem seriously", Nakamura says. Both say that they wanted to see the centre rebooted under new leadership and with a new name. Kishi says he wanted a "reborn CDB" — and in his view, the restructuring has met that aim. "Nothing's really changed," says Kishi.

But that is not how it looks to those on the ground. Hamada and other CDB researchers worry about whether they will be able to attract new principal investigators and postdoctoral researchers to the now-tarnished centre. It has lost two of its most prominent researchers in Sasai and Niwa, who is moving to Kumamoto

Yoshiki Sasai (right), co-author of the controversial stem-cell papers, bows after speaking at a press conference in April 2014.

University this year. Fujiwara has stayed at the CDB, but says that his research has suffered. When he was supported by secure funding from the centre, he had planned a five-year project to map proteins in the extracellular matrix of hair follicles. Now that he must secure year-to-year grants, that project looks harder to realize.

The impact of the STAP episode has also been felt further afield. Since August, policies for preventing and dealing with science misconduct have been released by the science ministry, the health ministry and other bodies. As part of this, data-management rules are being tightened to ensure that data are checked more often and are made available for verification. Such rules are frustrating scientists, and there are widespread concerns that research is becoming overly bureaucratic. "We had a pretty good system, but now we have to check this and that," says Kawai. "I don't want to grab time from scientists."

The new science-ministry guidelines make an institution responsible for fraud, and say that its funding can be cut if fraud-prevention measures are deemed inadequate. Imai says that this could be counterproductive, because it could push institutes and scientists to cover up suspected misconduct.

STAP could have a more subtle but pervasive impact on science, says Ito. Historically, young researchers in Japan have been able to take an experiment in a different direction from their supervisor's, and to follow their hunches down often productive paths. But the fear of fraud is likely to lead to stricter internal auditing, including the checking of lab notebooks, so "now they'll be afraid to. That will affect their ability to be innovative and they will be less motivated," says Ito. "That will be the biggest impact going forward." It is ironic, she says, that the experience of the CDB, which was born out of the desire to encourage young scientists, could end up stifling them.

Steneck says that it is useful to assess whether the STAP case was handled in the best way. He says that RIKEN responded well when the issue came to light: "They faced up to the problem quickly and brought in other opinions." But he questions the make-up of the Kishi committee, which was dominated by scientists. He thinks that a misconduct investigation should be led by experts in studying behaviour in scientific and research institutions. "Having a committee of non-experts gather their own evidence does not work."

Paul Taylor, director of the Office for Research Ethics and Integrity at the University of Melbourne in Australia, says that a Japanese investigatory organization akin to the US Office of Research Integrity might have helped. "It provides an independent source of expertise in conducting investigations, and perhaps ensures that trust in research is maintained," he says. He adds that Japan's tightened requirements on data management and research-integrity training might help to prevent scientific misconduct in the future — but there are limits. "If someone is able to justify to themselves why it is OK to fabricate an image, or steal someone's work, then they will do that even in the presence of the best research-integrity environment."

Hamada is optimistic that the CDB can move on. He hopes to negotiate with the science ministry to restore researchers' budgets within three years, and plans to continue the CDB's mission of gambling on promising young scientists. "I have to reorganize, refresh," he says. "My job is to forget about what happened."

That will not be easy for Fujiwara. He was just starting to get good data — but is worried that he will have trouble getting funding for the mouse experiments that he needs to even publish his work so far. "It was going to be an important year for us," he says. Now, he just hopes that his science can survive. ■

David Cyranoski *is a reporter for* Nature *based in Shanghai, China.*

1. Obokata, H. *et al.* Nature **505,** 641–647 (2014).
2. Obokata, H. *et al.* Nature **505,** 676–680 (2014).

A 2013 protest against the Syrian government pays tribute to those who died in explosions at the University of Aleppo earlier that year.

After the
ARAB SPRING

*Four years after revolutions shook governments in North Africa
and the Middle East, scientists face an uncertain future.*

BY MOHAMMED YAHIA AND DECLAN BUTLER

As a young physicist in Tunisia, Imen Sfar gave little thought to politics until March 2010, when a street vendor set himself on fire to protest about corruption in the city of Monastir, where Sfar worked.

Two months later, an influential blogger and close friend of Sfar's was arrested and interrogated for helping to organize a demonstration opposing Internet censorship in Tunisia. His detention gave Sfar the "courage factor", she says, to join the revolution that erupted in December 2010, after another

street vendor set himself aflame in Sidi Bouzid to protest against harassment by local officials. That act unleashed years of pent-up frustrations against the repressive government of President Zine al-Abidine Ben Ali, who had ruled the country for almost a quarter of a century (see *Nature* **469,** 453–454; 2011). "Events unfolded very quickly, with people taking to social media and the streets," recalls Sfar. "It was what I call the abolition of fear; we no longer were scared of anything."

The uprising triggered a wave of unrest that swept across North Africa and the Arabian

Peninsula. Demonstrations and revolts destabilized long-standing regimes, ultimately toppling rulers in Tunisia, Libya, Egypt and Yemen.

Nearly five years after the Arab Spring, the political situation in most of those countries remains volatile and many scientists are struggling more than they did before. This week, *Nature* profiles how researchers' situations have changed in Tunisia, Egypt and Syria, which have followed starkly different political trajectories.

In Tunisia, scientists are celebrating the

country's successful transition to democracy, although the hard-won political freedoms have not yet translated into changes in the research system. In nearby Egypt, many of the freedoms won through the revolution have been reversed and the status of scientists has improved little. Syria has deteriorated the most, spiralling into an all-out civil war that has driven most scientists to flee the country. Yet even amid the chaos, some researchers have remained in Syria, where they struggle to teach students and carry on their research in any way possible.

"I do not fear death, it will happen anywhere you go — but many students in Syria need our help," says Ahmad Almansour, a materials engineer at the University of Aleppo.

TUNISIA

Before the Arab Spring, Tunisia had a long tradition of support for education and research, unlike many of its neighbours. It produced more papers, relative to the size of its population, than any other Arab nation apart from Saudi Arabia, and it was the only one to invest more than 1% of its gross domestic product (GDP) in research and development.

But the repressive government kept tight control of university policies and stifled academic freedom, especially in disciplines that delve into potentially controversial territory, such as the social sciences. Faouzia Charfi, a retired physicist and veteran opponent of the regime's human-rights restrictions, tells how researchers needed permission for almost everything, even organizing academic conferences and collaborating with foreign peers.

The regime feared people acting on their own initiative, says Charfi, who became junior minister for higher education in the first post-revolution transition government. Universities and researchers had little freedom to develop their own policies or strategies. And regime bureaucrats blocked attempts to build links between universities and industry, stifling the kind of innovation that builds economies and creates jobs, she says.

The police state established by Ben Ali affected most aspects of everyday life. Few dared to criticize the government or even to discuss politics, and people could never be sure that neighbours or friends were not part of the regime's web of informers, says Sfar. Instead, they took refuge in the safe subject of sport. "Before the revolution, Tunisians discussed nothing but football," she says.

After Ben Ali fled Tunisia in January 2011, freedom of expression exploded overnight, says Sfar. But the euphoria soon dimmed as the country went through four years of successive governments and political turmoil, including the assassination of leading opposition politicians. At times, many feared that the democratic gains of the revolution would be reversed. But Tunisia's robust civil society

Physicist Imen Sfar is optimistic that newfound freedoms will translate into better research in Tunisia.

kept the revolution on track, with huge street demonstrations and strikes. After the moderate Islamist party Ennahda reached a compromise with secular parties, Tunisia last year became the first and only Arab Spring country to successfully transition to democracy.

In a historic moment, parliament overwhelmingly passed a constitution that guaranteed free speech, freedom of expression, religious freedom and equality between the sexes. The constitution also explicitly protects scientific and academic freedom, and mandates that the state "supply the means necessary for the development of scientific research and technology". In October, a secularist party won the parliamentary elections, and in December its candidate Beji Caid Essebsi was elected as the nation's president.

The biggest change since the revolution, says Sfar, "is this kind of feeling of freedom to say what we think about politics and the administration without fearing serious repercussions". At the professional level, she says that the revolution has had little impact on her field of condensed-matter physics and she has been preoccupied with starting her career and a family. She has taken a position at the University of Tunis El Manar, and is contemplating switching to theoretical work on materials with unusual electrical and magnetic properties because she lacks access to the kind of equipment she used while completing her PhD in France.

More broadly, the dramatic changes in Tunisian society have yet to permeate the research and higher-education systems. The past quarter of a century of authoritarian rule has left a system that is far from realizing its potential, says Hechmi Louzir, director of the Pasteur Institute in Tunis. The regime's tight grip on free thinking stifled creativity, he says.

"People didn't stray from the beaten path," says Louzir. Sfar says that even now, many of her students do not seem keen to think for themselves, and they expect lecturers "to show them one, and only one, road to the truth".

And Charfi is disappointed that researchers have not yet taken advantage of their new freedoms to bring in sweeping reforms, attributing the reticence to the highly conservative culture of the country's scientific community.

Educational reform and innovation in science are important, say researchers, because they will help Tunisia to create wealth and jobs — a priority given the country's dismal economic state and high unemployment. And Tunisians are ever mindful of local threats; just last month radical Islamist militants killed 22 people in a terrorist attack on the Bardo National Museum in Tunis.

Still, Tunisian researchers point to some positive changes. In the past, political nepotism and influence often determined who held key university posts, but democratic elections now decide, they say.

As Sfar looks back over the changes, she sees grounds for optimism. "There is now freedom of expression and thought, and an extraordinarily active civil society," she says.

EGYPT

After a citizen uprising toppled the regime of President Hosni Mubarak in February 2011, microbiologist Ramy Aziz returned to Egypt full of hope for change.

"I had wanted to come back to Egypt to bring the scientific experience I gained in bioinformatics back to Cairo University," says Aziz, who was at San Diego State University in California at the time. "The revolution was the spark that changed everything overnight."

Students, researchers and professors took part in the uprising and their expectations soared in the months that followed. The government promised to increase science funding and many expatriate researchers vowed to return home.

Researchers called for reforming outdated policies — such as requirements that slowed down the transfer of samples across borders or limited interactions with industry. Such changes, they said, would help Egypt to target its most pressing problems: water and energy insecurity, poverty and mounting unemployment.

But the high hopes of Aziz and others have yet to be fulfilled. The government approved a doubling in science funding in 2012, but there was no clear agenda on how to use those funds and the research ministry failed to spend more than 80% of its budget. "On the ground we did not feel any change in the budget really," says Aziz, who is now at Cairo University.

Then, on 3 July 2013, a popularly backed coup removed the first democratically elected president, Mohamed Morsi, who represented the Muslim Brotherhood. Several of the gains that researchers won after the uprising have since been reversed. In June last year, for example, newly elected president Abdel Fattah al-Sisi scrapped the law introduced after the uprising that allowed professors to choose their university's leaders. The positions are now appointed by the president (see *Nature* **511**, 5; 2014).

"From a scientific point of view, I can't see any change in Egypt," says Aziz. He does feel that university students and postgraduates have become bolder and more willing to challenge their professors and ask for their rights — a positive outcome. But he does not expect that to make a big difference in Egypt's science.

The biggest win for scientists is constitutional, says Alaa-Eldin Adris, who chairs the department of petroleum and energy engineering at the American University in Cairo. The new constitution, passed in January 2014, states that 1% of GDP should be spent on scientific research, and that this can be increased up to international standards. "On paper this looks good," Adris says. But

he cautions that the constitutional change is only the first step. "It is worthless unless it creates laws that are enforced."

According to the World Bank, science spending in Egypt rose from 0.24% of GDP in 2009 before the uprising to 0.43% in 2011. But most of the increase goes to salaries rather than to funding research. In the *Global Competitiveness Report 2014–2015*, produced by the World Economic Forum, Egypt ranks among the worst ten countries in the world in the quality of its scientific research institutes.

"We have many researchers, which is good, but when it comes to using science to solve problems or drive economic growth, we are almost at zero," says Adris.

Aziz says that working life in Egypt today is a continuous battle in a stilted system, where he lacks access to core facilities and an environment that supports research. In Egypt, he focuses on teaching and mentoring postgraduates, but he returns to the United States for a few months a year to conduct most of his experiments. When he works there, primers for DNA synthesis arrive a day after he orders them. "In Egypt, you get primers in three weeks. If they don't work well, you will need another three weeks," he says.

Aziz thinks that the current political leadership is sincere in its goals to support science but has done little to actually help researchers. Others, however, accuse the regime of taking a strong stand against freedom of expression within universities. Students have been arrested during protests on the campuses of Cairo's Al-Azhar University and at Cairo University and security forces have frequently stormed the university gates, leading to violent confrontations with protesters.

Many young people who might otherwise have helped to build up Egyptian science are being driven out. Islam Kotb left after he lost many friends in a violent dispersal of protesters in August 2013. "Leaving the country became my ultimate aim at that time and I accepted the first job offer I got," he says. Now working as a critical-care specialist in Jeddah, Saudi Arabia, he hopes to return to Egypt one day, but stresses "that scientific research can never succeed and flourish in such an environment".

For Aziz and Adris, the real threats to research in Egypt are the bureaucracy (buying chemicals can take several months, for example) and the government's failure to engage in serious reform, such as overhauling education in ailing public schools and universities. None of these problems has been addressed by any of the successive ruling regimes, they say.

"I personally expect that by 2020 there will be little change, if any, in scientific research and output," says Aziz.

Yet even with all these problems, he still

Ramy Aziz had hoped that the ousting of President Mubarak would make science easier to do in Egypt.

finds hope. "I see excellent students trying to learn against insurmountable odds," he says. "You need to create a new generation with a different mindset and proper education that will eventually replace the current people in control, with their outdated vision."

SYRIA

In January 2013, students were gathering for their mid-year examinations at the University of Aleppo when a series of rockets hit residence halls and other buildings. Smoke billowed across the campus and people ran in panic to find shelter. The attack killed at least 82 people, including students and refugees who had sought protection at the university. The rebels and the government blamed each other for the bombing.

Materials engineer Almansour was lucky that day. The blast blew out the windows of the building where he was working, 300 metres from the centre of the explosion, but he and his family were unharmed. Although most Syrian scientists have fled the country to escape the ongoing civil war, Almansour feels an obligation to his students
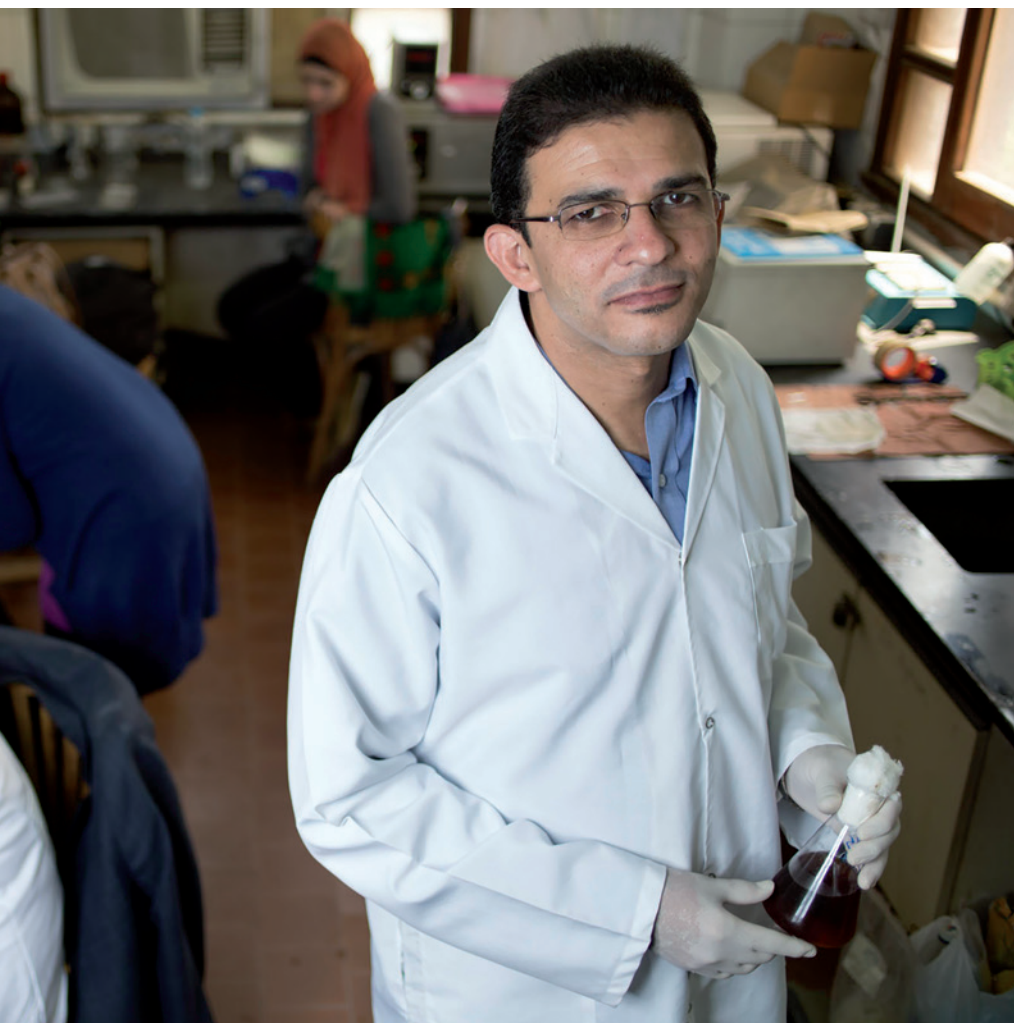
"THE REVOLUTION WAS THE SPARK THAT CHANGED EVERYTHING OVERNIGHT."

and is staying put. "If we all leave the country, who is going to teach them?" he asks.

Whereas Tunisia and Egypt have started to stabilize politically, the situation in Syria has grown steadily worse. What started as an uprising against lack of freedom and an oppressive autocratic regime has turned into a civil war with many factions and little hope of resolution.

Throughout the four years of turmoil, however, some education and research have continued. When Michel Rahal graduated in 2013 from Damascus University with a bachelor's in applied chemistry, he had hoped to pursue a postgraduate degree in Europe. But the government eliminated almost all scholarships for overseas education, and in any case, Rahal felt a responsibility to stay in his country during these challenging times. He is now studying for a master's degree at the Higher Institute of Applied Science and Technology in Damascus, although he feels that his options are limited while in Syria.

Some of those who have left say that it was impossible for them to remain in the country. Ahmad Salman, who asked to have his real name changed for fear of retribution against his family still in Syria, is one of those who tried to stay as long as possible. "Despite the difficulties of life in Syria during the war with the lack of security, access to water or fuel and the huge inflation, I lived with the hope that the revolution would end in favour of the people," he says. But when the regime passed a law drafting all men into the army, he left his position at Al-Baath University and fled to Turkey, where he is now jobless. "I only had two choices, either join the military and be on the front lines to kill or be killed by my countrymen, or get out of Syria."

Most international researchers in Syria have also fled the violence, and that has impaired research in the country. The International Center for Agricultural Research in the Dry Areas (ICARDA), one of the biggest research centres in Syria, struggled for almost two years to keep its doors open before leaving the country for neighbouring Lebanon in late 2012. Mahmoud Solh, a geneticist and director-general of ICARDA, says that looters had repeatedly attacked the facilities and stolen vehicles, computers and other equipment. But the centre managed to save its agricultural gene bank, one of the most important such collections, before it relocated.

Researchers who remain in Syria say that the working conditions there are extremely difficult. "Students in the medical and engineering school need equipment and materials that they cannot get," says Zeina Al-Ahmad, a finance specialist at Tishreen University in Latakia. "We cannot participate in conferences as we used to either, which has also affected the quality of research in Syria." The high inflation rate, which peaked at 121% in August 2013, has hit everyone hard, making it nearly impossible for scientists to fund their research and publishing activities.

And security remains a major concern for students and faculty members, who must often pass through dangerous war zones to reach their universities. Almansour lives close to his university's campus, but many students have had to drop out for fear of their safety. "The number of students attending classes is about 35–40% of the registered number," he says. "I am most worried about the possible destruction of the infrastructure. It took us a long time to build campuses, but if the fighting spares universities and schools then we will be able to go on in spite of the terrible circumstances."

As bad as conditions are in Aleppo and other areas controlled by the government, the situation can be even worse for those elsewhere. The extremist group ISIS has taken control of large areas of eastern Syria, where it has closed schools and universities and is forcing teachers to work in its newly founded schools or lose their lives, impelling many professors and teachers to flee. There are few signs that things are likely to improve, with ISIS and other groups now involved in the civil war.

Given all the problems, Rahal plans to move to Europe as soon as he can. Many other researchers are also leaving, says al-Ahmad. "The young generation is trying to finish university and find a job abroad." ∎

Mohammed Yahia is chief editor of Nature *Middle East.* **Declan Butler** is a senior reporter for Nature *in France.*

COMMENT

ILLUSTRATION BY GREG CLARKE

Time for one-person trials

Precision medicine requires a different type of clinical trial that focuses on individual, not average, responses to therapy, says **Nicholas J. Schork**.

Every day, millions of people are taking medications that will not help them. The top ten highest-grossing drugs in the United States help between 1 in 25 and 1 in 4 of the people who take them (see 'Imprecision medicine'). For some drugs, such as statins — routinely used to lower cholesterol — as few as 1 in 50 may benefit[1]. There are even drugs that are harmful to certain ethnic groups because of the bias towards white Western participants in classical clinical trials[2].

Recognition that physicians need to take individual variability into account is driving huge interest in 'precision' medicine. In January, US President Barack Obama announced a US$215-million national Precision Medicine Initiative. This includes, among other things, the establishment of a national database of the genetic and other data of one million people in the United States.

Classical clinical trials harvest a handful of measurements from thousands of people. Precision medicine requires different ways of testing interventions. Researchers need to probe the myriad factors — genetic and environmental, among others — that shape a person's response to a particular treatment.

Studies that focus on a single person — known as *N*-of-1 trials — will be a crucial part of the mix. Physicians have long done these in an ad hoc way. For instance, a doctor

may prescribe one drug for hypertension and monitor its effect on a person's blood pressure before trying a different one. But few clinicians or researchers have formalized this approach into well-designed trials — usually just a handful of measurements are taken, and only during treatment.

If enough data are collected over a sufficiently long time, and appropriate control interventions are used, the trial participant can be confidently identified as a responder or non-responder to a treatment. Aggregated results of many *N*-of-1 trials (all carried out in the same way) will offer information about how to better treat subsets of the population or even the population at large. ▶

▶ Formalizing and scaling up the *N*-of-1 approach means solving various practical problems. These include exploiting the diversity of health-monitoring devices, developing new ones and identifying appropriate disease biomarkers, such as tumour DNA circulating in the bloodstream. It will also require a cultural shift on many levels — in regulatory agencies, in pharmaceutical companies and, most of all, in the clinic.

A WORLD OF DIFFERENCE

Discovering that an intervention works well in certain groups happens relatively rarely and often by chance. Researchers typically get disappointing results with a drug in large, population-based trials. This leads them to conduct ad hoc post-trial analyses, to try to identify the factors that cause some of the people in the trial to seem to be responsive[3].

For instance, the drug Gleevec (imatinib) was found to double survival rates of leukaemia patients[4] with a chromosomal abnormality in their tumours called the Philadelphia translocation. Similarly, it turns out that Erbitux (cetuximab) improves the survival of people with colorectal cancer whose tumour cells carry a mutated *EGFR* gene but not a mutated *KRAS* gene[5].
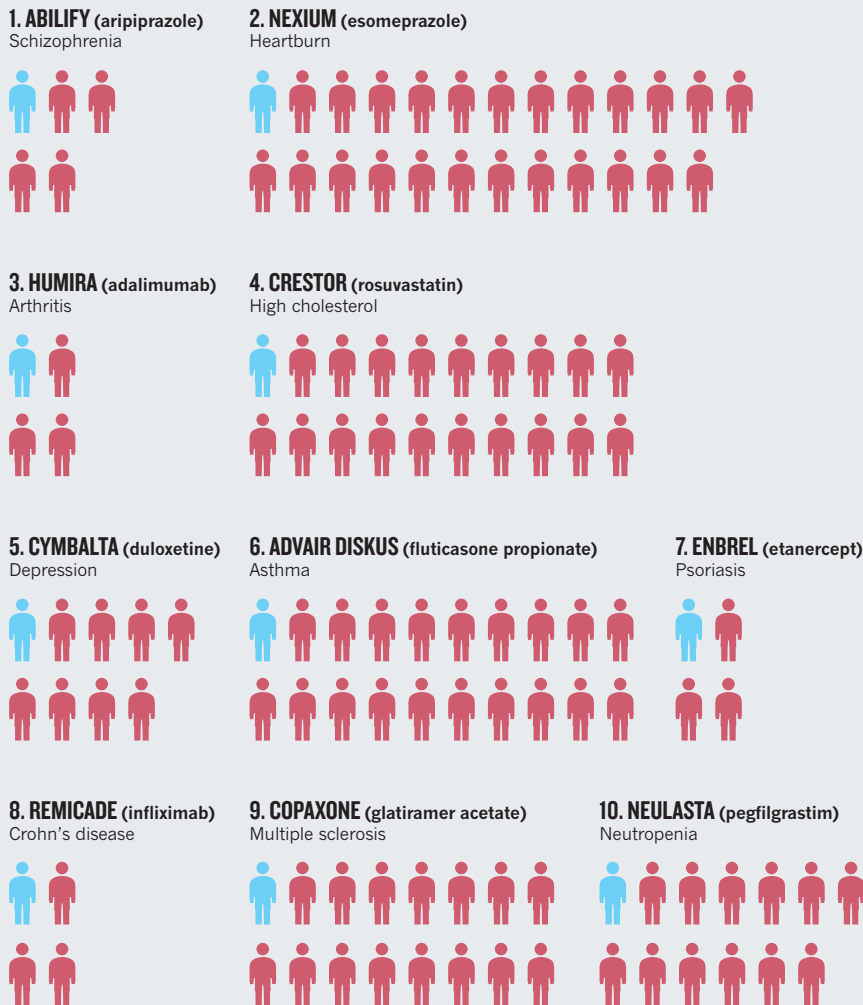
This approach to discovery is inefficient at best. Conventional phase III trials involve thousands of people. The intervention being tested is often given at random to one group while another group receives a sham treatment, such as a sugar pill or the standard treatment that physicians would give such patients. Because scant data are collected on factors such as genetics, lifestyles and diets, the results of these trials often indicate the need for yet another study to validate the effectiveness of the intervention among the apparent responders and to establish the underlying mechanisms.

Various trial designs have been developed that better account for variability between patients. Basket trials, which have mainly been used for cancer, test the effectiveness of an intervention on the basis of its mode of action, regardless of what disease it was designed to treat. For instance, the US National Cancer Institute's MATCH Trial, expected to launch in May, will use genetic markers from tumours to assign 1,000 people who have different types of late-stage cancer to different treatment 'baskets'. Baskets will include around 30 patients, all of whom share certain genetic anomalies; participants will be given one of around 25 drugs. The effectiveness of the different matches of treatment to genetics will be compared as the trial unfolds.

In umbrella trials, researchers test the effectiveness of multiple drugs in a study of a single disease. This approach is being used in the ongoing Stand Up To Cancer and Melanoma Research Alliance Dream Team clinical trial, in which I am involved, run by Yale University

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (aripiprazole)
Schizophrenia

2. NEXIUM (esomeprazole)
Heartburn

3. HUMIRA (adalimumab)
Arthritis

4. CRESTOR (rosuvastatin)
High cholesterol

5. CYMBALTA (duloxetine)
Depression

6. ADVAIR DISKUS (fluticasone propionate)
Asthma

7. ENBREL (etanercept)
Psoriasis

8. REMICADE (infliximab)
Crohn's disease

9. COPAXONE (glatiramer acetate)
Multiple sclerosis

10. NEULASTA (pegfilgrastim)
Neutropenia

Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr78f.

in New Haven, Connecticut, and the Translational Genomics Research Institute in Phoenix, Arizona. We are giving genomically guided treatment to around 70 people out of 100 with late-stage melanoma — choosing from an array of 40 or so drugs (the rest of the participants are receiving the standard treatment usually given to such patients). Here, the effectiveness of matches between drugs and genetics will be compared with standard care. Lastly, adaptive trials aim to match interventions to patients while the study is ongoing, on the basis of patient responses.

Even these trial designs may not be personalized enough. Among people who share, say, a particular mutation known to be targeted by a specific drug, many other factors can contribute to any one person's responsiveness. This is particularly true for those with cancer. The drug vemurafenib, for instance, was approved in the United States to treat late-stage melanoma in people whose tumours

carry the BRAF(V600E) mutation. But some tumour cells develop other anomalies that make them resistant to the drug. Thus clinicians considering whether to prescribe vemurafenib may need to take into account a whole slew of biomarkers.

COMPARING TREATMENTS

In *N*-of-1 trials, all sorts of relevant data will need to be collected for one person, as frequently as possible — perhaps every day or periodically over months or years. The usual design and statistical safeguards could be employed, such as blinding patients and experimenters to the drugs being tested, and the use of control interventions (such as periods of standard care). In addition, appropriate crossover designs, in which different interventions are administered to the same person alternately (possibly with 'wash-out' periods in between to allow the drugs' effects to wear off) would enable experimenters to

compare the effect of different treatments in the same person.

If done properly, claims about a person's response to an intervention could be just as well supported by a statistical analysis as by analyses designed to assess population-level responses on the basis of classical clinical trials[6].

An example of this approach is a study[7] in Australia, which measured reported pain levels, swelling and other symptoms associated with osteoarthritis and chronic pain in 132 people taking different drugs over three years. For each person, measurements were taken every 2 weeks for 12-week periods, when the patient was either off or on a particular drug. By comparing the data collected before and after the different treatments, the researchers showed that, although initially costly, the formalized N-of-1 trials resulted in more-effective prescriptions.

Sometimes N-of-1 trials will be neither appropriate nor feasible. For instance, the costs are probably too high for public-health studies that investigate the effect of a population-wide intervention, such as adding fluoride to drinking water. Making objective claims about individual responses requires taking appropriate measures (of tumour progression, say) repeatedly and efficiently. Yet it is not always clear what to measure. Only a fraction of the thousands of proposed biomarkers have been shown to be useful in the clinic.

But in many instances, an N-of-1 approach is ideal. Such studies are already being done for some rare diseases by necessity, but often without the use of sophisticated trial designs and without necessarily collecting the appropriate information to make hypotheses about the drug's mechanism. Many experimental drugs are also administered in 'compassionate use' settings. And many widely used drugs are provided to combat diseases for which they were not approved ('off label' prescription), for people who fail to respond to all other treatments. Examples include uses of the drug mexiletine to treat the rare muscle disease non-dystrophic myotonia, and experimental treatments for the Ebola virus.

Well-designed N-of-1 trials could also be useful in the early stages of clinical drug development or in repurposing drugs — for exploring the molecular and physiological effects of a new compound (or of an old compound in a new context). Likewise, studies investigating the safety and appropriate dosages of drugs could take an N-of-1 approach. Currently, phase I and II clinical trials usually involve giving different amounts of a drug to a small group of healthy volunteers. Better would be to tailor dosages to individuals' metabolic profiles.

N-of-1 trials could be designed to guide clinicians in detecting disease onset. For instance, US physicians generally view levels of a blood protein called CA125 greater than 30 or 35 as an indication of ovarian cancer. However, a level of 20 or 25 may be a cause for concern if the person's average CA125 levels hovered around 10 or 15 over the previous year[8]. Establishing personal thresholds for uncovering disease onset is the goal of the registered clinical trial known as the Tanner Project (www.tannerproject.org), in which I am involved.

By looking for commonalities across multiple N-of-1 studies — in which the same types of data are collected using the same procedures — researchers should be able to draw inferences about the effectiveness of an intervention in certain subsets of the population, such as in people sharing particular genetic features, as well as in the whole population.

"A key component will be transforming everyday clinical care into solid N-of-1 trials."

Various teams are developing and testing algorithms to match interventions, or a combination thereof, to individuals on the basis of their genetic make-up, biochemistry, diet and other factors. For instance, matching drugs to tumour profiles is a key goal of the Stand Up To Cancer umbrella trial.

MAKING IT HAPPEN

There are significant barriers to making N-of-1 trials commonplace. Regulatory agencies, researchers and clinicians are rightfully wary of moving away from classical clinical trials. Pharmaceutical companies tend to focus on drugs that are likely to be used by thousands or millions of people. What is more, tailoring treatments to patients is costly. For example, the cancer-care company Foundation Medicine in Cambridge, Massachusetts, charges patients between US$5,000 and $7,500 to sequence their tumours and to use the results to advise on treatments. And there is a lot of work to be done on biomarkers, monitoring devices, study designs and data-analysis methods.

A key component will be transforming everyday clinical care into solid N-of-1 trials. In my view, the time is ripe for three reasons.

First, there is a growing interest in 'omics' assays that expose people's unique characteristics at the molecular level. Researchers and clinicians are assaying people's blood metabolites (their metabolome) and the microbes in their bodies (their microbiome) as well as their DNA and RNA[9]. Second, cheap and efficient devices that collect health data are becoming available, such as the Apple Watch, continuous glucose monitors and portable electroencephalogram (EEG) monitors. Lastly, governments and life-sciences funding bodies worldwide are increasingly supporting a more targeted approach as well as patient engagement in medicine, such as through the US Patient-Centered Outcomes Research Institute, established in 2010.

I am confident that, ultimately, governments, regulatory agencies and pharmaceutical companies will support sophisticated, well-designed N-of-1 trials. These could save the millions of US dollars that are spent on inappropriate interventions, the management and treatment of persistent or recurring diseases, and on conventional phase III trials (which can cost between $100 million and $700 million per drug). Regulatory agencies such as the US Food and Drug Administration are beginning to recognize the importance of individual responses[10]. And sufficient financial or market incentives provided by governments could persuade pharmaceutical companies to broaden their focus away from 'blockbuster' drugs — especially given the poor rates of return on drug discovery in recent years.

Key to making precision medicine mainstream is the ongoing shift in the relationship between patients and physicians. A major advantage of the N-of-1 approach over classical trials is that patients are no longer guinea pigs, whose involvement in a study may help only future generations. In N-of-1 trials, the effectiveness of different treatments are vetted for the actual participants. Indeed, members of hundreds of patient-advocacy groups, for instance for rare genetic diseases, are eager to be involved in studies to test candidate drugs.

Physicians are having to become more acutely aware of the unique circumstance of each patient — something most people have long called for. ∎

Nicholas J. Schork *is director of human biology at the J. Craig Venter Institute in La Jolla, California, USA. He is also professor at the University of California, San Diego, and at the Translational Genomics Research Institute (TGen) in Phoenix, Arizona, USA.*
e-mail: nschork@jcvi.org

1. Mukherjee, D. & Topol, E. J. *Prog. Cardiovasc. Dis.* **44,** 479–498 (2002).
2. Currie, G. P., Lee, D. K. & Lipworth, B. J. *Drug Saf.* **29,** 647–656 (2006).
3. Uryniak, T. *et al. Stat. Biopharmaceut. Res.* **3,** 476–487 (2011).
4. Druker, B. J. *et al. N. Engl. J. Med.* **344,** 1038–1042 (2001).
5. Karapetis, C. S. *et al. N. Engl. J. Med.* **359,** 1757–1765 (2008).
6. Kravitz, R. L. & Duan, N. *Design and Implementation of N-of-1 Trials: A User's Guide* (Agency for Healthcare Research and Quality, 2014).
7. Schuffham, P. A. *et al. J. Gen. Intern. Med.* **25,** 906–913 (2010).
8. Drescher, C. W. *et al. J. Clin. Oncol.* **31,** 387–392 (2013).
9. Chen, R. *et al. Cell* **148,** 1293–1307 (2012).
10. US Food and Drug Administration. *Paving the Way for Personalized Medicine: FDA's Role in a New Era of Medical Product Development* (US Department of Health And Human Services, FDA, 2013).

P values are just the tip of the iceberg

Ridding science of shoddy statistics will require scrutiny of every step, not merely the last one, say **Jeffrey T. Leek** and **Roger D. Peng**.

There is no statistic more maligned than the *P* value. Hundreds of papers and blogposts have been written about what some statisticians deride as 'null hypothesis significance testing' (NHST; see, for example, go.nature.com/pfvgqe). NHST deems whether the results of a data analysis are important on the basis of whether a summary statistic (such as a *P* value) has crossed a threshold. Given the discourse, it is no surprise that some hailed as a victory the banning of NHST methods (and all of statistical inference) in the journal *Basic and Applied Social Psychology* in February[1].

Such a ban will in fact have scant effect on the quality of published science. There are many stages to the design and analysis of a successful study (see 'Data pipeline'). The last of these steps is the calculation of an inferential statistic such as a *P* value, and the application of a 'decision rule' to it (for example, $P < 0.05$). In practice, decisions that are made earlier in data analysis have a much greater impact on results — from experimental design to batch effects, lack of adjustment for confounding factors, or simple measurement error. Arbitrary levels of statistical significance can be achieved by changing the ways in which data are cleaned, summarized or modelled[2].
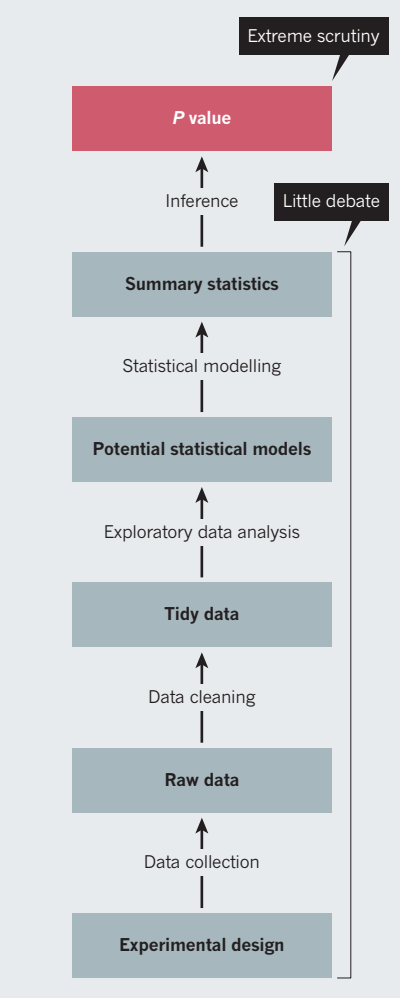
P values are an easy target: being widely used, they are widely abused. But, in practice, deregulating statistical significance opens the door to even more ways to game statistics — intentionally or unintentionally — to get a result. Replacing *P* values with Bayes factors or another statistic is ultimately about choosing a different trade-off of true positives and false positives. Arguing about the *P* value is like focusing on a single misspelling, rather than on the faulty logic of a sentence.

Better education is a start. Just as anyone who does DNA sequencing or remote-sensing has to be trained to use a machine, so too anyone who analyses data must be trained in the relevant software and concepts. Even investigators who supervise data analysis should be required by their funding agencies and institutions to complete training in understanding the outputs and potential problems with an analysis.

There are online courses specifically

DATA PIPELINE

The design and analysis of a successful study has many stages, all of which need policing.

Extreme scrutiny

P value ← Inference ← Little debate

Summary statistics

Statistical modelling

Potential statistical models

Exploratory data analysis

Tidy data

Data cleaning

Raw data

Data collection

Experimental design

designed to address this crisis. For example, the Data Science Specialization, offered by Johns Hopkins University in Baltimore, Maryland, and Data Carpentry, can easily be integrated into training and research. It is increasingly possible to learn to use the computing tools relevant to specific disciplines — training in Bioconductor, Galaxy and Python is included in Johns Hopkins' Genomic Data Science Specialization, for instance.

But education is not enough. Data

analysis is taught through an apprenticeship model, and different disciplines develop their own analysis subcultures. Decisions are based on cultural conventions in specific communities rather than on empirical evidence. For example, economists call data measured over time 'panel data', to which they frequently apply mixed-effects models. Biomedical scientists refer to the same type of data structure as 'longitudinal data', and often go at it with generalized estimating equations.

Statistical research largely focuses on mathematical statistics, to the exclusion of the behaviour and processes involved in data analysis. To solve this deeper problem, we must study how people perform data analysis in the real world. What sets them up for success, and what for failure? Controlled experiments have been done in visualization[3] and risk interpretation[4] to evaluate how humans perceive and interact with data and statistics. More recently, we and others have been studying the entire analysis pipeline. We found, for example, that recently trained data analysts do not know how to infer *P* values from plots of data[5], but they can learn to do so with practice.

The ultimate goal is evidence-based data analysis[6]. This is analogous to evidence-based medicine, in which physicians are encouraged to use only treatments for which efficacy has been proved in controlled trials. Statisticians and the people they teach and collaborate with need to stop arguing about *P* values, and prevent the rest of the iceberg from sinking science. ∎

Jeffrey T. Leek *and* **Roger D. Peng** *are associate professors of biostatistics at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, USA.*
e-mail: jleek@jhsph.edu

1. Trafimow, D. & Marks, M. *Basic Appl. Soc. Psych.* **37**, 1–2 (2015).
2. Simmons, J. P., Nelson, L. D. & Simonsohn, U. *Psychol. Sci.* **22**, 1359–1366 (2011).
3. Cleveland, W. S. & McGill, R. *Science* **229**, 828–833 (1985).
4. Kahneman, D. & Tversky, A. *Econometrica* **47**, 263–291 (1979).
5. Fisher, A., Anderson, G. B., Peng, R. & Leek, J. *PeerJ* **2**, e589 (2014).
6. Leek, J. T. & Peng, R. D. *Proc. Natl Acad. Sci. USA* **112**, 1645–1646 (2015).

SPRING BOOKS

CLIMATE ECONOMICS

The high road

Michael Grubb is both swept away and frustrated by Nicholas Stern's argument for tackling climate change.

Nicholas Stern's *Why Are We Waiting?* is an important but frustrating book. It is important because it embodies the thinking of probably the most influential person at the nexus of climate change, economics and international policy. It is frustrating because it does not answer the titular question. It presents compelling economic logic for urgent action on climate change, but fails to get to the heart of what stands in the way.

Stern, who was chief economist at the World Bank in the early 2000s, produced the hugely influential 2006 *Stern Review on the Economics of Climate Change* while head of the UK Government Economic Service. A crucial achievement of the report was Stern's framework, which bridged 15 years of apparent impasses between the dominant scientific and economic perspectives. It offered a cost–benefit structure that agreed with the scientists, that strong action was needed to counter anthropogenic global warming. It ignited debate on climate change throughout the economics community, which until then had considered the subject of marginal interest. Nine years on, Stern concludes in *Why Are*

We Waiting? that the case for action is now even more urgent.

The book is astonishingly broad in scope and vision. Its core argument is that the next two decades will be decisive, as a critical period for emissions coincides with massive structural change in the world economy. These are conjoined challenges, and Stern has long argued that "high-carbon growth is likely to self-destruct". How the economic changes are channelled will determine the choice "between peril and prosperity".

Stern's coverage of the science is taut and displays his eye for relevant dimensions of the big picture, linked to human realities. He writes, for instance, that unchecked greenhouse-gas emissions are likely to cause carbon dioxide concentrations to reach 750 parts per million of the atmosphere by 2100, resulting
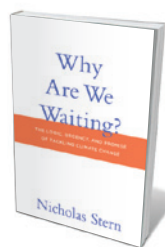
Why Are We Waiting?: The Logic, Urgency, and Promise of Tackling Climate Change
NICHOLAS STERN
MIT Press: 2015.

in a median temperature increase of 4 °C or more above preindustrial levels, which would transform Earth's "physical and human geography". So "deserts, coastlines, rivers, rainfall patterns — the reasons we live where we do — would be redrawn". *Homo sapiens*, he says, "has not experienced anything like this". Stern notes that the uneven distribution of temperature increases could "correspond to average land temperatures as much as 4–10 °C above the preindustrial", and could trigger frequent summer heatwaves as severe as the Russian disaster of 2010, which killed 55,000 people.

Stern endorses the goal of checking emissions by enough to allow a 50:50 chance of staying below a 2 °C increase, but the window for that is closing rapidly. Far from being a burden justified by the risks, however, he sees the endeavour as "very attractive". The centrepiece of his argument — drawing heavily on last year's New Climate Economy report, which he co-chaired — is the potential for a tsunami of innovation to engulf the fossil-fuel economy, galvanizing economic growth.

However, Stern does not adequately delve into the fact that too much analysis, modelling and policy is still dominated by the idea that there is a natural 'least-cost' energy system, and that market forces would deliver it if we only corrected market failures. This reflects the inbuilt assumptions of neoclassical economics. But the broad sweep of economic processes and thought spans at least two other domains. One concerns the behavioural and organizational realities that underpin the huge potential to improve energy efficiency, and other 'easy wins'. The other is evolutionary economics: understanding how complex

systems evolve and can be locked into trajectories by incumbent industries, unless strong pressures force a change in course. These strands of economics are crucial to understanding the energy–climate nexus and policies for transformation. The economics mainstream is still largely in thrall to the idea that competitive markets drive innovation, but liberalization of the energy sector destroyed UK energy research and development. The energy literature explains why, but Stern hardly touches on this and thus misses a chance to help to educate fellow economists.

Stern's real speciality is displayed most in his discussion of the ethics of how to weigh costs and benefits over time, as well as broader issues in moral and political philosophy. The first is technical, but is at the heart of the post-Stern review debate: prevalent economic tools 'discount' almost anything more than a few decades ahead. Critics often cite philosopher David Hume to support arguments that we should weight impacts on future generations on the basis of 'revealed preference' in markets. To these critics, Stern cites Hume on the need for "governors and rulers" to overcome the "impatience" of individuals. Stern maps out his case with a precision reminiscent of philosopher and mathematician Bertrand Russell. He demolishes economists who, trying to avoid explicit ethical debates, forget the ethical assumptions of their discipline.

Towards the end, Stern distils his thoughts on the international climate scene. Although he overlooks some key respects in which a binding treaty — the Kyoto Protocol — was crucial in forming European climate policy,

huge value remains in his wisdom and his experience of international debates.

However, Stern's coverage of lessons from economic history and theory is messier. It meanders through market failures, policy uncertainties, the European Union's emissions-trading system, the debate over reserves of fossil fuels that must be left unburnt to satisfy emissions targets, shale gas, the German nuclear phase-out, failed Chinese dams and more. There are trenchant points about six market failures and the six areas of policy required to address them, but little structure behind the tales, and no compelling narrative on how policy failures might have been overcome. Instead, Stern embarks on a demolition of his critics. The integrated assessment models beloved of many technical analysts come in for particular ire. Stern views them as "simplistic attempts to shoehorn the deep and dynamic issues into inappropriate or narrow models", noting that many "can be profoundly misleading", and evoking John Maynard Keynes' old punchline: "It is better to be roughly right than precisely wrong."

Most seriously, Stern has misdiagnosed the major obstacles to climate action as a failure of policy-makers to understand the scale of the risks, to grasp that delay is dangerous or to appreciate the feasibility of and potential for low-carbon paths. These are not the root of the problem, and Stern never gets to it, despite providing much of the intellectual material.

Most policy-makers accept the need for action. Ignorance, inattention and ideology have all played a part in the waiting game, but it also springs from microrealities of politics

and economics. Stern is a big thinker, used to the broad sweeps of economic development and global issues. But the crux is local people and businesses. Voters might like the idea of clean energy, but oppose wind farms next door; back emission reductions and profess support for market-based solutions, but oppose increased energy prices. There is little on energy prices in *Why Are We Waiting?*, but climate policy in Brussels and Washington DC is concerned with little else.

The real challenge in controlling climate change is many-layered, but must include attempts to align global need for transformation with the old adage that all politics is local. A brief reference to border measures, to deter industries from migrating to escape carbon controls, signals a shift in Stern's thinking; but the argument is not developed further.

Stern deserves much credit for moving the economics debate on so far in a decade. But *Why Are We Waiting?* is a depressing reminder of the sheer size of the elephant. The complexities of climate change are beyond the ken of any individual, even such a big thinker as Stern. Developing the institutions, intellectual frameworks and impact channels required for effective 'solution science', matched to the sheer range and scale of the problems and opportunities, remains very much a work in progress. ∎

Michael Grubb *is professor of international energy policy and climate change at University College London, editor-in-chief of* Climate Policy, *and author of* Planetary Economics. *e-mail: m.grubb@ucl.ac.uk*

We are the 98%

Nathaniel Comfort unpicks the metaphors in a trio of books exploring the 'junk'-ridden genome.

The language of DNA is a veritable cornucopia of metaphor and cliché. Since James Watson and Francis Crick solved the double helix, biologists have imagined DNA as an information-storage device: magnetic tape, a computer program or, most commonly, a book that contains the instructions for making a cell's proteins. In multicellular organisms, this precious tome is secured in the vault of the nucleus, the membrane of which isolates and protects nature from nurture.

But if a genome is text, it is badly edited. Most DNA is gibberish, full of stutters, snippets of doggerel from other species, and echoes of quiescent viruses. In humans, only

about 2% of the genome encodes proteins. Much — but not all — of the remaining 98% is evolutionary detritus. In the 1960s, researchers learned that non-coding DNA can serve vital functions, such as regulating gene action and building ribosomes. The remainder they began to call junk.

Today, junk DNA is at the heart of the most radical transformation of how we understand the genome since the information metaphor. Three books — *The Deeper Genome* by John Parrington, *Junk DNA* by Nessa Carey and *Biocode* by Dawn Field and Neil Davies — present a vision of the twenty-first-century genome. Their relative success hinges on metaphor and imagery, both in how they

Biocode: The New Age of Genomics
DAWN FIELD AND NEIL DAVIES
Oxford Univ. Press: 2015.

The Deeper Genome: Why There Is More to the Human Genome Than Meets the Eye
JOHN PARRINGTON
Oxford Univ. Press: 2015.

Junk DNA: A Journey Through the Dark Matter of the Genome
NESSA CAREY
Icon: 2015.

conceive the genome and in the writing itself.

In September 2012, the ENCODE (Encyclopedia of DNA Elements) consortium announced that its multi-year international effort to catalogue the various types of DNA sequence had assigned "biochemical function" to 80% of the human genome. Incautious reporters began shouting that junk was bunk, even though scientific consensus maintains that most genomes contain large amounts of it. The subsequent debate upregulated public interest in non-coding DNA — but how do we talk about DNA now?

The title *Biocode* forestalls any doubt ▶

systems evolve and can be locked into trajectories by incumbent industries, unless strong pressures force a change in course. These strands of economics are crucial to understanding the energy–climate nexus and policies for transformation. The economics mainstream is still largely in thrall to the idea that competitive markets drive innovation, but liberalization of the energy sector destroyed UK energy research and development. The energy literature explains why, but Stern hardly touches on this and thus misses a chance to help to educate fellow economists.

Stern's real speciality is displayed most in his discussion of the ethics of how to weigh costs and benefits over time, as well as broader issues in moral and political philosophy. The first is technical, but is at the heart of the post-Stern review debate: prevalent economic tools 'discount' almost anything more than a few decades ahead. Critics often cite philosopher David Hume to support arguments that we should weight impacts on future generations on the basis of 'revealed preference' in markets. To these critics, Stern cites Hume on the need for "governors and rulers" to overcome the "impatience" of individuals. Stern maps out his case with a precision reminiscent of philosopher and mathematician Bertrand Russell. He demolishes economists who, trying to avoid explicit ethical debates, forget the ethical assumptions of their discipline.

Towards the end, Stern distils his thoughts on the international climate scene. Although he overlooks some key respects in which a binding treaty — the Kyoto Protocol — was crucial in forming European climate policy,

huge value remains in his wisdom and his experience of international debates.

However, Stern's coverage of lessons from economic history and theory is messier. It meanders through market failures, policy uncertainties, the European Union's emissions-trading system, the debate over reserves of fossil fuels that must be left unburnt to satisfy emissions targets, shale gas, the German nuclear phase-out, failed Chinese dams and more. There are trenchant points about six market failures and the six areas of policy required to address them, but little structure behind the tales, and no compelling narrative on how policy failures might have been overcome. Instead, Stern embarks on a demolition of his critics. The integrated assessment models beloved of many technical analysts come in for particular ire. Stern views them as "simplistic attempts to shoehorn the deep and dynamic issues into inappropriate or narrow models", noting that many "can be profoundly misleading", and evoking John Maynard Keynes' old punchline: "It is better to be roughly right than precisely wrong."

Most seriously, Stern has misdiagnosed the major obstacles to climate action as a failure of policy-makers to understand the scale of the risks, to grasp that delay is dangerous or to appreciate the feasibility of and potential for low-carbon paths. These are not the root of the problem, and Stern never gets to it, despite providing much of the intellectual material.

Most policy-makers accept the need for action. Ignorance, inattention and ideology have all played a part in the waiting game, but it also springs from microrealities of politics

and economics. Stern is a big thinker, used to the broad sweeps of economic development and global issues. But the crux is local people and businesses. Voters might like the idea of clean energy, but oppose wind farms next door; back emission reductions and profess support for market-based solutions, but oppose increased energy prices. There is little on energy prices in *Why Are We Waiting?*, but climate policy in Brussels and Washington DC is concerned with little else.

The real challenge in controlling climate change is many-layered, but must include attempts to align global need for transformation with the old adage that all politics is local. A brief reference to border measures, to deter industries from migrating to escape carbon controls, signals a shift in Stern's thinking; but the argument is not developed further.

Stern deserves much credit for moving the economics debate on so far in a decade. But *Why Are We Waiting?* is a depressing reminder of the sheer size of the elephant. The complexities of climate change are beyond the ken of any individual, even such a big thinker as Stern. Developing the institutions, intellectual frameworks and impact channels required for effective 'solution science', matched to the sheer range and scale of the problems and opportunities, remains very much a work in progress. ∎

Michael Grubb *is professor of international energy policy and climate change at University College London, editor-in-chief of* Climate Policy, *and author of* Planetary Economics. *e-mail: m.grubb@ucl.ac.uk*

We are the 98%

Nathaniel Comfort unpicks the metaphors in a trio of books exploring the 'junk'-ridden genome.

The language of DNA is a veritable cornucopia of metaphor and cliché. Since James Watson and Francis Crick solved the double helix, biologists have imagined DNA as an information-storage device: magnetic tape, a computer program or, most commonly, a book that contains the instructions for making a cell's proteins. In multicellular organisms, this precious tome is secured in the vault of the nucleus, the membrane of which isolates and protects nature from nurture.

But if a genome is text, it is badly edited. Most DNA is gibberish, full of stutters, snippets of doggerel from other species, and echoes of quiescent viruses. In humans, only

about 2% of the genome encodes proteins. Much — but not all — of the remaining 98% is evolutionary detritus. In the 1960s, researchers learned that non-coding DNA can serve vital functions, such as regulating gene action and building ribosomes. The remainder they began to call junk.

Today, junk DNA is at the heart of the most radical transformation of how we understand the genome since the information metaphor. Three books — *The Deeper Genome* by John Parrington, *Junk DNA* by Nessa Carey and *Biocode* by Dawn Field and Neil Davies — present a vision of the twenty-first-century genome. Their relative success hinges on metaphor and imagery, both in how they

Biocode: The New Age of Genomics
DAWN FIELD AND NEIL DAVIES
Oxford Univ. Press: 2015.

The Deeper Genome: Why There Is More to the Human Genome Than Meets the Eye
JOHN PARRINGTON
Oxford Univ. Press: 2015.

Junk DNA: A Journey Through the Dark Matter of the Genome
NESSA CAREY
Icon: 2015.

conceive the genome and in the writing itself.

In September 2012, the ENCODE (Encyclopedia of DNA Elements) consortium announced that its multi-year international effort to catalogue the various types of DNA sequence had assigned "biochemical function" to 80% of the human genome. Incautious reporters began shouting that junk was bunk, even though scientific consensus maintains that most genomes contain large amounts of it. The subsequent debate upregulated public interest in non-coding DNA — but how do we talk about DNA now?

The title *Biocode* forestalls any doubt ▶

that the authors hew to the information metaphor. Breathlessly, Field and Davies survey the greatest hits and promises of genomics, including *Jurassic Park*-style reanimation of extinct species, the microbiome and environmental engineering. The thin chapters blurt out strings of recent findings, each capped with a crescendo of sensational speculations that mostly rehearse familiar ethical questions. Critical distance is achieved with the time-honoured double negative: "Might a lawyer one day argue that deliberately *not* giving [our children] the best genes available is a form of abuse? It is not inconceivable to imagine a world where natural reproduction would seem primitive and even barbaric." It concludes by exhorting us to set our sights on a global genome project to understand "the software that shapes our living planet". The biocode is Gaia plus DNA. But two clichés do not make a right. *Biocode* simply extends the text metaphor to the macrocosm.

The old metaphor is not wrong; it is incomplete. In the new genome, lines of static code have become a three-dimensional tangle of vital string, constantly folding and rearranging itself, responsive to outside input. The roots of this idea run deep. In her 1983 Nobel lecture, geneticist Barbara McClintock called the genome a "sensitive organ of the cell". McClintock, who discovered mobile genetic elements in the 1940s, had named them controlling elements because she thought they composed the regulatory system that governed gene action. In 1980, Ford Doolittle and Carmen Sapienza proposed that transposons were molecular parasites, jumping into genomes to propagate themselves. Parasitic transposons are now textbook knowledge, but McClintock's larger point holds: the genome is dynamic, full of regulatory elements that respond to environmental cues.

The Deeper Genome is the only book of the three that credits McClintock as a progenitor of the three-dimensional genome. A scientist and journalist, Parrington covered the ENCODE story for *The Times* in 2012; his book enriches those accounts with historical and scientific context. The science is better than the history. He provides a fine discussion of recent support for McClintock's often-overlooked late work on how stress can activate transposition, but he perpetuates the

myth that at first no one thought transposition was real. The contested point was actually McClintock's interpretation of mobile elements as controllers of gene action. Parrington's strongest chapters survey the emerging view of gene regulation, including DNA folding, epigenetics and regulatory RNA. Overall, this is a faithful, engaging portrait of the twenty-first-century genome.

Finally, *Junk DNA*, like the genome, is crammed with repetitious elements and superfluous text. Bite-sized chapters parade gee-whizz moments of genomics. Carey's *The Epigenetics Revolution* (Columbia University Press, 2012) offered lucid science writing and vivid imagery. Here the metaphors have been deregulated: they metastasize through an otherwise knowledgeable survey of noncoding DNA. At one point, the reader must run a gauntlet of baseball bats, iron discs, Velcro and "pretty fabric flowers" to understand "what happens when women make eggs". The genome seems to provoke overheated prose,

unbridled speculation and Panglossian optimism. Junk DNA produces a lot of DNA junk.

The idea that the many functions of noncoding DNA make the concept of junk DNA obsolete oversells a body of research that is exciting enough. ENCODE's claim of 80% functionality strikes many in the genome community as better marketing than science.

Still, as with McClintock, the larger point holds: the genome is more than a set of rules and parts descriptions. Finding apt imagery to replace the dead metaphor of the 'instruction book of life' could enable us to break free of the cliché of nature versus nurture. It could usher in a more democratic conception of life, in which all the world's a cell, and all the genes and genomes merely players. ∎

Nathaniel Comfort *is professor of the history of medicine at Johns Hopkins University in Baltimore, Maryland. He is working on a biography of DNA.*
Twitter: @nccomfort

NEW IN PAPERBACK

Highlights of this season's releases
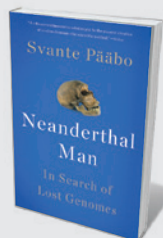
Neanderthal Man: In Search of Lost Genomes
Svante Pääbo BASIC *2015*
Pioneer of ancient-DNA studies Svante Pääbo was inspired in his youth by ancient Egyptian history. Feeling that this field moved too slowly, he decided to study medicine instead, and went on to sequence the first full Neanderthal genome in 2010. Here he details the technicalities of his life's work and the incremental discoveries, such as genetic intimations that modern humans and Neanderthals had mixed, which generated our theories of human evolution. (See Henry Gee's review: *Nature* **506,** 30–31; 2014.)

An improbable journey

Adrian Woolfson enjoys two studies on microbial life's trek towards complexity.

In 1676, the Dutch merchant and amateur scientist Antoni van Leeuwenhoek submitted an essay to the Royal Society of London detailing a singular discovery. This was the world of unicellular organisms, which he observed using a self-designed microscope. Three hundred years later, Leeuwenhoek's "animalcules" were shown to hold the secret to the evolution of complex life on Earth.

In his imaginative and beautifully written *The Vital Question*, evolutionary biochemist Nick Lane defines a genealogy that links the descendants of the Cambrian explosion — the first appearance of morphologically complex animals in the fossil record, about 540 million years ago — to the simple organisms that preceded them. In so doing, he persuades us that comprehending the structure, function, behaviour, genetics and evolution of microorganisms is necessary for a deep understanding of complex life, and of the processes that undermine it, including diseases and ageing. This visceral insight into the largely uncharted expanses of microbial existence could also form the basis of a predictive science enabling us to speculate about the nature of potential life on other planets.

Biophysicist Paul Falkowski's entertaining, easy-to-read and historically rich *Life's Engines*, meanwhile, uses the work of microbiologist Carl Woese to trace complex life back to its three lines of descent: bacteria, archaea and eukaryotes. By studying the RNA sequences of ribosomes — the cellular machines that make proteins — Woese was able to show that Charles Darwin was correct in suggesting that all life arose from a single, now-extinct, common ancestor.

It remains unclear how and when life first originated on Earth, but we know that the first unicellular organism emerged between 3.6 billion and 2.7 billion years ago, giving rise to bacteria and archaea, which have no nucleus or other sub-cellular organelles. The evolutionary engine of life then seems to have got stuck, idling along at the unicellular level for another 2 billion to 3 billion years. Falkowski explains how unicellular organisms, although morphologically challenged, managed to perfect the basic biochemical 'engines' that would power all forms of life on Earth. According to Lane, the stagnation occurred because the molecular motors that drive the biochemistry of bacteria and archaea were unable to cross the energetic threshold necessary for the evolution of complex form. This energetic

THE EVOLUTIONARY ENGINE **GOT STUCK,** IDLING ALONG AT THE **UNICELLULAR LEVEL.**

constraint on life is the central focus of *The Vital Question*.

It derives, Lane explains, from two principal design features that all living things use to power themselves. The first is the use of high-energy molecules of ATP, the chemical currency of energy transfer. The second is the idiosyncratic 'chemiosmotic' force, which moves protons and facilitates the continuous generation of ATP. Both Lane and Falkowski describe these molecular processes compellingly. Although adequate to power single bacteria-sized cells, the method constrains the allowable surface-to-volume ratio of a living cell. Lane argues, however, that around 1.5 billion years ago this energetic constraint was overcome by an improbable endosymbiosis event: an

The Vital Question: Why is Life the Way it is?
NICK LANE
Profile: 2015.

Life's Engines: How Microbes Made Earth Habitable
PAUL G. FALKOWSKI
Princeton Univ. Press: 2015.

ancestral archaean host engulfed a small population of symbiotic bacteria, resulting in the first eukaryotic cell, the forebear of complex life.

Lane recounts how over time, the engulfed bacteria jettisoned most of their genes that were unrelated to energy production; these were either lost permanently or relocated to the cell nucleus. There they continued to fulfil their original functions, or formed the raw material for the evolution of new genes with unexpected roles, such as transcription factors — proteins that bind to DNA. This allowed embryonic stem cells to be patterned in three-dimensional space. What remained of the imbibed bacteria, with their pared-down genomes and surrounding membranes, became energy-generating mitochondria. The acquisition of these organelles enabled eukaryotic cells to expand their volume by up to 15,000 times that of the average bacterium, and to support a genome around 5,000 times larger. Lane's important realization is that this also gifted eukaryotic cells with about 200,000 times more energy per gene than the average prokaryotic cell. This over-cranking of the evolutionary engine allowed for the development of a baroque diversity in the nature and extent of cellular gene and protein expression.

Although readily accommodated by classic Darwinian evolutionary theory, the horizontal, sudden and co-operative nature of Lane's evolutionary narrative differs from the incremental, vertical and competitive ▶
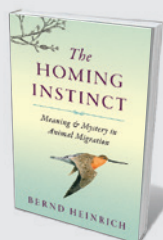
The Homing Instinct: Meaning and Mystery in Animal Migration
Bernd Heinrich MARINER *2015*
Erudite naturalist Bernd Heinrich attributes the instinct for migration to an affinity for 'home', from beavers' skilful dam-building to the joyful dance of Alaskan cranes returned to their nesting pond. (See Joel Greenberg's review: *Nature* **508,** 317; 2014.)

The Accidental Species: Misunderstandings of Human Evolution
Henry Gee UNIV. CHICAGO UNIVERSITY PRESS *2015*
Nature's palaeontology editor, Henry Gee, condemns the idea that our species is the pinnacle of evolution, arguing that traits prized as uniquely human, such as creativity, are not. (See Tim Radford's review: *Nature* **503,** 34–35; 2013.)

▶ features of the more canonical one. As elegant as the details underpinning the thesis seem, it is occasionally hard to distinguish between fact and speculation. Lane has, nevertheless, made a bold and commendable attempt to sketch out a highly challenging scientific issue for a general audience. In so doing, he has reaffirmed the importance of a largely overlooked area of basic research, and has generated testable hypotheses about the origins of complex life.

Falkowski covers some of the same details of the evolution of microbes and their contribution to complexity, including the historical origins of the concept of endosymbiosis. However, his focus is primarily on how microorganisms have made Earth habitable, perhaps most notably with the development of oxygen-generating photosynthesis by cyanobacteria. This leads him to touch on humanity's potential to undermine Earth's systems.

History has shown how modifications to microbial biochemistry affect global geophysical processes. For example, following 200 million to 300 million years of photosynthesis by ancient microorganisms, oxygen became a significant component of Earth's atmosphere, increasing ozone levels, reducing the greenhouse effects of gases such as methane and leading to one of the most extensive glaciations in the planet's history. Humanity's interference with natural biological processes risks damaging Earth in ways that cannot be predicted.

What is clear is that a deep understanding of how complex life originated will provide insights into human biology and the nature of disease processes. It may also enable the generation of forms of life unconstrained by the contingent processes that locked life into its current trajectory. Life as we know it may eventually be supplanted, perhaps one day even being viewed as a primordial soup that facilitated the emergence of silicon-based existence. ∎

Adrian Woolfson *is the author of* Life Without Genes.
e-mail: adrianwoolfson@yahoo.com

Journal of the plague years

Mark Dybul applauds the latest chapter in an account of a life at the leading edge of HIV research and policy.

Virologist Peter Piot's *AIDS Between Science and Politics* is a terrific follow-up to his highly acclaimed memoir *No Time to Lose* (W. W. Norton, 2012). It demonstrates the deep intellectual lessons of a lifetime at the cutting edge of science and politics. Piot's narrative ranges from his thrilling, on-the-ground experiences in remote regions of Africa as a young scientist and member of the team that identified Ebola, to the high-altitude reflections of his years as executive director of the Joint United Nations Programme on HIV/AIDS (UNAIDS).

Three messages underpin the book's nine chapters. One is that the HIV epidemic generated an unprecedented local and global response, recast many development and health paradigms, and ultimately triggered treatments that have saved millions of lives. The second is that progress was made only when various scientific disciplines, on-the-ground implementation strategies and politics were aligned. And the third? That AIDS is not over.

This Changes Everything: Capitalism vs the Climate
Naomi Klein ALLEN LANE *2015*
Unafraid to name and shame fossil-fuel junkies hooked on a billion-dollar industry, Naomi Klein investigates capitalism and climate change. She sees the global crisis as a potential spur to positive action, as happened with the women's rights movement. (See Nico Stehr's review: *Nature* **513,** 312; 2014.)
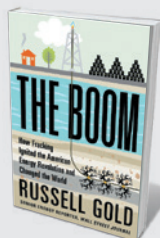
The Boom: How Fracking Ignited the American Energy Revolution and Changed the World
Russell Gold SIMON AND SCHUSTER *2015*
Journalist Russell Gold traces the rise of fracking, a tale of innovation and investment — such as ex-oilman Aubrey McClendon's 260,000 land acquisitions in Texas's Barnett Shale. (See Chris Nelder's review: *Nature* **508,** 185; 2014.)

Piot observes that HIV is one of the most devastating plagues in history, reducing life expectancy in countries such as Botswana by around 30% — and wiping out decades of gains in health and development. Even in places that have not been affected on that national scale, it has devastated groups of the most marginalized and vulnerable people, including men who have sex with men, transgender people, people who inject drugs, sex workers, prisoners and others. Piot also points out the links with historical injustices such as the 1948–94 apartheid in South Africa, where the social fabric was so badly damaged and the mistrust of the authorities grew so deep that it exacerbated the outbreak of the epidemic.

Piot traces the early days of the response to HIV/AIDS, when governments closed their ears and wallets to the growing epidemic.

AIDS Between Science and Politics
PETER PIOT
Columbia Univ. Press: 2015.

People with the virus rallied to fight for attention and resources, with increasing energy and sophistication. Remarkable individual and community action effected a breathtaking shift from a paternalistic, government-only approach to development and health, to one in which partnership and inclusivity provided a more effective response.

From such a heady climate were born the Global Fund to Fight HIV, Tuberculosis and Malaria in 2002 and the US President's Emergency Plan for AIDS Relief (PEPFAR) in 2003, both aiming to provide funding for programmes in countries most affected by the diseases. Perhaps of even greater significance was the growth of multisectoral local institutions, including national AIDS councils in dozens of countries, often reporting to the head of state. And there has been an increase of billions of dollars in domestic financing for HIV/AIDS response. As a result, many millions of lives have been saved and lifted up.

The theme of effective partnerships that span disciplines, politics and theologies (religious, scientific and otherwise) permeates the book, and is reflected in its title. The science is complex, as are the politics,

and they are intimately linked owing to social elements, and the discrimination and stigma they engender — what Piot calls the challenge of "sex, drugs and rock and roll". Data, modelling and advocacy were essential to move national and global politics, but political leadership also drove a demand for scientific advance, data and results.

At a national level, Piot praises former president of Botswana Festus Mogae for his personal leadership in declaring, in 2001, that HIV was a threat to his country's existence. Mogae dedicated significant national resources to antiretroviral therapy at a time when much of the global health community scoffed at the viability of such a programme in Africa. These moves were strengthened by an alliance with the Bill & Melinda Gates Foundation and with pharmaceutical giant Merck, creating what Piot deems perhaps the most successful and impactful public–private partnership in history. At the global level, Piot singles out the push by then-UN secretary-general Kofi Annan for the creation of the Global Fund, and George W. Bush's efforts towards the inauguration of PEPFAR.

Piot is honest about failures. He cites an effort he pioneered in the mid-2000s to engage the US and European pharmaceutical industries in expanding access to antiretroviral treatment for people in low-income countries. The opposition came from a diverse group of people and institutions, including academics and political activists, who feared that participation by the pharmaceutical industry would be geared towards increasing profits, not expanding access.

He is also clear about the need to remain focused and vigilant. In his view, the epidemic could be significantly reduced with sufficient resources, channelled to where they would make the most difference. This is the heavy baton that has been picked up by Piot's successor at UNAIDS, Michel Sidibé.

AIDS Between Science and Politics is a must-read for anyone interested in the HIV/AIDS epidemic. More broadly, it offers lessons — and interesting anecdotes — useful in the response to Ebola and indeed to every challenge in global health and development. ∎

Mark Dybul *is executive director of the Global Fund to Fight AIDS, Tuberculosis and Malaria in Geneva, Switzerland.*
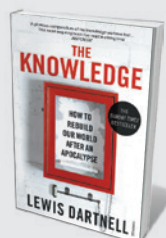
The Knowledge: How to Rebuild Our World After an Apocalypse
Lewis Dartnell VINTAGE *2015*
In a post-apocalyptic world, could we rebuild civilization? Lewis Dartnell condenses millennia of achievement into a handbook on mastering Earth's resources to produce food, energy and medicines using our "greatest invention": science.
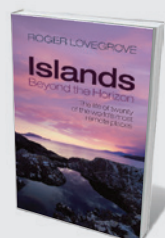
Islands Beyond the Horizon: The Life of Twenty of the World's Most Remote Places
Roger Lovegrove OXFORD UNIV. PRESS *2015*
In his profile of inaccessible islands, Roger Lovegrove's admiration for wildlife shines. But from Russia's ice-locked Wrangel Island, where polar bears banquet on walruses, to the Pacific Tuamotus, few such idylls remain pristine.

MEDICAL HISTORY

Pioneer of polio eradication

Tilli Tansey extols a biography of determined vaccine trailblazer Jonas Salk.

On 12 April 1955, across the United States, "church bells tolled, horns honked, and sirens rang" in celebration: the largest clinical trial ever undertaken had reported that the first polio vaccine was safe and effective. So writes Charlotte Jacobs in her riveting biography of the vaccine's discoverer, *Jonas Salk*. Two years earlier, poliomyelitis had killed or paralysed almost 36,000 US children. It has been estimated that before the vaccine there were 600,000 cases a year worldwide. Salk's triumph made him a global household name, and the gongs began rolling in.

As Jacobs shows, the tale of Salk's discovery is one of grind, intrigue, rivalry, politics and dirty tricks. Add commercial interests (pharmaceutical giant Eli Lilly made US$30 million from polio vaccine in 1955 alone) and Salk's extramarital entanglements while wedded to artist (and muse to Pablo Picasso) Françoise Gilot, and the mix becomes even headier. Jacobs contextualizes the polio effort with Salk's work on influenza, multiple sclerosis and HIV/AIDS — and with the Salk Institute for Biological Studies in La Jolla, California, which he created and directed, and from which he was ultimately excluded.

Growing up in the Bronx, New York, in a family of Russian Jewish immigrants, Salk was dominated by his ambitious mother. He came to crave doing things his own way — a
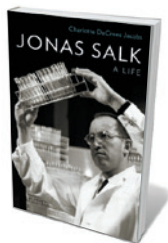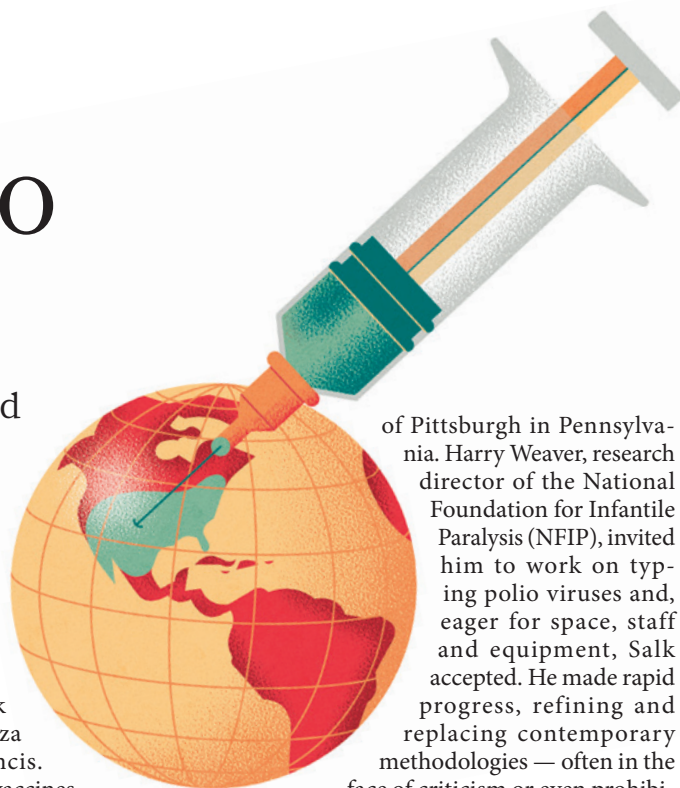
Jonas Salk: A Life
CHARLOTTE DECROES JACOBS
Oxford Univ. Press: 2015.

tenacity that shaped both his scientific success and his professional difficulties. Early on, at the New York University College of Medicine, Salk worked with influenza expert Thomas Francis. At the time, antiviral vaccines (against smallpox, rabies and yellow fever) used artificially weakened live virus. Salk and Francis developed an experimental technique that used killed virus to stimulate antibody production and confer immunity, suggesting a powerful therapeutic approach.

In 1941, Salk followed Francis to Michigan, pursuing the holy grail — an influenza vaccine. He modified laborious procedures to culture the virus and develop vaccine production, usually with only gloves and a mask as protection. He supervised clinical tests on patients at two psychiatric institutions, deliberately infecting some with influenza — a practice common until the Nuremberg Code of 1947 offered some protection to human research subjects — and in 1945 advised the US surgeon general to vaccinate 8 million soldiers. Without consulting Francis, Salk signed an exclusive contract with pharmaceutical firm Parke, Davis to provide details of production methods that he devised. This departure from academic etiquette did not go unnoticed; nor did his writing for nonprofessional publications such as *Parents Magazine*. Tensions grew as Francis received honours, while Salk was ignored.

In 1947, Jacobs recounts, Salk left to establish an influenza lab at the University of Pittsburgh in Pennsylvania. Harry Weaver, research director of the National Foundation for Infantile Paralysis (NFIP), invited him to work on typing polio viruses and, eager for space, staff and equipment, Salk accepted. He made rapid progress, refining and replacing contemporary methodologies — often in the face of criticism or even prohibition from the NFIP advisory committee. Weaver supported his protégé, and the foundation came to see Salk as its 'poster scientist', wheeled out for public and media events. Fellow researchers continued to carp.

Undeterred, Salk pioneered a killed-virus vaccine and organized safety testing and field trials. Those leading up to the 1955 announcement involved more than 1.5 million children, tens of thousands of doctors and nurses, and 220,000 volunteers. The clamour attendant on success (a film was mooted, to star Marlon Brando) laid Salk open to charges of self-aggrandisement.

Almost immediately, problems arose. Batches of vaccine became contaminated, and physicians inoculated family and friends while leaving first-grade children, the most vulnerable group, unprotected. Over a few months, 260 individuals contracted polio directly or indirectly from a single substandard preparation. Several states suspended vaccination. Massachusetts halted its programme when children had received only the first of three shots; in July 1955, some 4,000 contracted polio and 1,700 were paralysed. Salk was unfairly associated with these errors and was castigated, especially by scientific colleagues. However, tighter adherence to his

Water 4.0

David Sedlak YALE UNIV. PRESS *2015*
From Roman aqueducts to desalinization plants, David Sedlak's study overflows with facts about water management. Chlorine by-products could be carcinogens, so he argues that water treatment needs another upgrade. (See Margaret Catley-Carlson's review: *Nature* **505,** 288–289; 2014.)
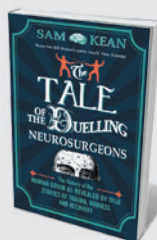
The Tale of the Duelling Neurosurgeons

Sam Kean BLACK SWAN *2015*
Crammed with curious anecdotes from neuroscience's gory past, Sam Kean's book ranges from the crude methods of early brain studies (including the beheading of criminals to use as test subjects) to the prion disease kuru, which spreads through cannibalism.

protocols and resumption of vaccination meant that six years after the vaccine was introduced, polio was almost eradicated in the United States.

Much of the scientific establishment closed ranks against Salk. He was given the prestigious Lasker Award for clinical medical research in 1956, but Swedish virologist Sven Gard dealt his Nobel nomination a fatal blow by sneering that the vaccine was a technical advance, not a discovery. Nor was Salk elected to the US National Academy of Sciences. Virologist Albert Sabin — bombastic, imperious and galled by Salk's success — continued to develop a live, orally delivered poliovirus preparation. By 1961, Sabin's vaccine had performed well in trials and the American Medical Association began to promote it. Salk's vaccine was, for a time, superseded, and his efforts to improve its potency stymied.

Salk moved on, although he remained involved with the polio vaccine. Influenced by chemist C. P. Snow's 1959 book *The Two Cultures and the Scientific Revolution*, he launched a research institute integrating social responsibility and the humanities with the biological sciences. The Salk Institute recruited some of the great biologists of the time, including Jacob Bronowski, Francis Crick and Jacques Monod. But Salk was unable to translate his lofty ideals into practical management. His research from the 1960s onwards, on immune responses in cancer, multiple sclerosis and, later, HIV/AIDS, met with ambivalence. He was increasingly derided by the very scientists whom he had recruited.

In many ways, Salk was ahead of his time, notably in public engagement and in his multidisciplinary agenda. A polio vaccine would have emerged without him, but it was his vision and willpower that produced the first, and a descendant of it is still the basis of many public-health programmes. Yet universal polio eradication remains a dream: cases continue to appear in Pakistan, Afghanistan and Nigeria, and have resurged in recent years in Syria. ∎

Tilli Tansey *is professor of the history of modern medical sciences at Queen Mary, University of London.*
e-mail: t.tansey@qmul.ac.uk
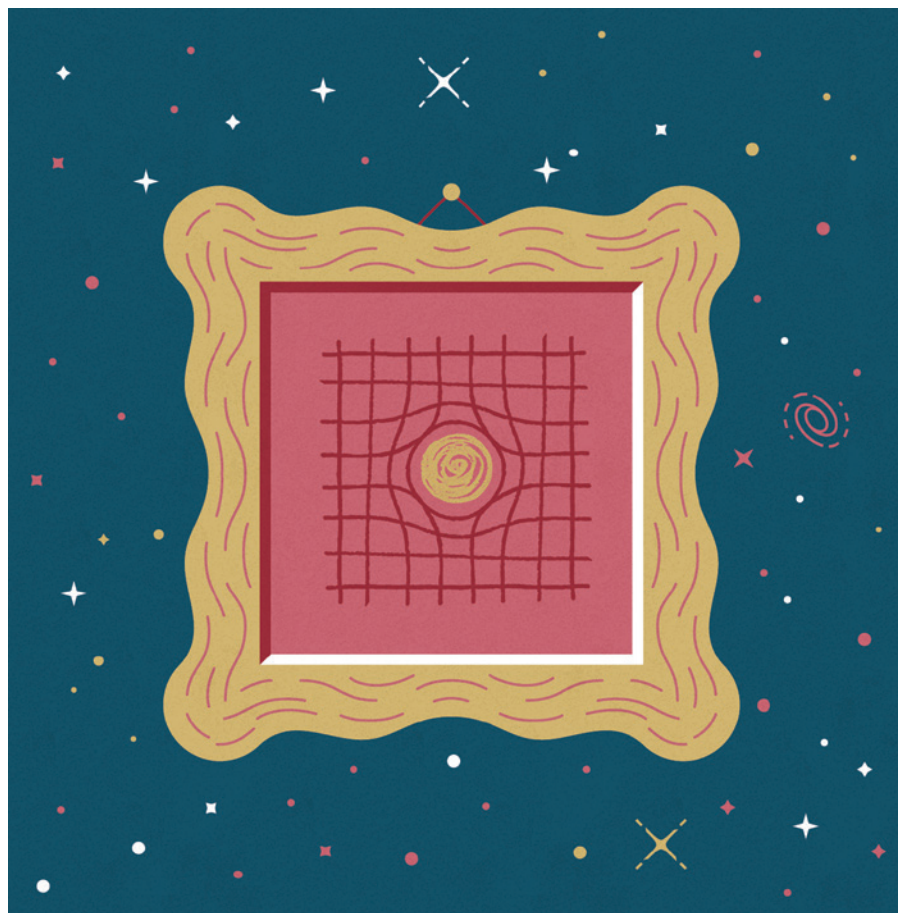

PHYSICS

One hundred years of general relativity

Pedro Ferreira looks back at how Einstein himself and a panoply of other physicists have framed the theory.

Until very recently, relativists were few and often self-taught. General relativity still had the stigma of being esoteric, pointless and, well, hard. In some places you could find specialized graduate courses, but on the whole, if you were at all interested in expanding universes and black holes, you were left to your own devices. That is what happened to me.

I studied engineering and did not enjoy it very much. But during the course on electromagnetism, I discovered Albert Einstein's world of special relativity. ▶
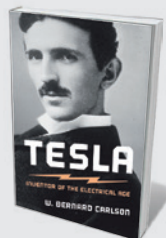
Tesla: Inventor of the Electrical Age
W. Bernard Carlson PRINCETON UNIV. PRESS *2015*
Over-hyped eccentric or electricity wizard? Bernard Carlson's account of Nikola Tesla's life at the turn of the twentieth century recalls the inventor's great creations, such as the alternating-current motor, as well as the unfulfilled promise of wireless power. (See Patrick McCray's review: *Nature* **497,** 562–563; 2013.)
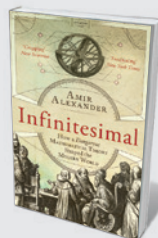
Infinitesimal: How a Dangerous Mathematical Theory Shaped the Modern World
Amir Alexander ONEWORLD *2015*
Through religious and revolutionary figures of the seventeenth century, Amir Alexander tells the history of the struggle for mathematics' place in society. The 'heretical' concept of infinitesimals, the indivisible points of a line, takes centre stage.

The mathematics was seductive, the paradoxes were mind-blowing, and it set me up to try to learn his general theory of relativity. That theory explains how the gravitational force is nothing more than space-time bending and warping as it responds to the presence of energy and mass. To understand this revolutionary viewpoint, I had to look for the right book, something that could lead me through all the intricacies of Riemannian geometry, which overturns the rules of Euclidean geometry that we learn in school. Yet I also needed to understand the physics: the bending of light and the orbit of Mercury.

One book stood out: Einstein's popular *Relativity: The Special and the General Theory*. Published in German in 1916, following Einstein's groundbreaking 1915 paper on the general theory of relativity, it was translated into English in 1920. In 2015, we see the publication of a special anniversary edition, as well as an annotated version of the manuscript of the paper in Hanoch Gutfreund and Jürgen Renn's *The Road to Relativity*.

Einstein's book *Relativity* was supposed to be understandable by all, yet to have enough maths to allow the more educated reader to get into the guts of his ideas. It has very few equations, rendering it less explanatory and more illustrative. But there are definitely a lot of words. Einstein set himself the task of explaining the concepts and ideas behind his theory, using situations from everyday life, such as trains moving on platforms or clocks on walls. His prose is tempered with some philosophical considerations, for example a discussion of the 'a priori' assumption that empty space exists.

Dare I say it, I found the prose inelegant. This caught me by surprise. I had read some of Einstein's 1905 papers, including the one introducing special relativity, and had thought them gems. *Relativity*, by contrast, was not particularly clear and a bit dull. Einstein had declared, in the introduction, that he would repeat himself frequently, "without paying the slightest attention to the elegance of the presentation". In this he might have been following the dictum of physicist Ludwig Boltzmann, who pinned

Relativity: The Special and the General Theory (100th Anniversary edition)
ALBERT EINSTEIN
Princeton Univ. Press: 2015

The Road to Relativity: The History and Meaning of Einstein's "The Foundation of General Relativity" Featuring the Original Manuscript of Einstein's Masterpiece
HANOCH GUTFREUND AND JÜRGEN RENN
Princeton Univ. Press: 2015

down the concept of entropy and declared that "matters of elegance should be left to the tailor and to the cobbler". Nevertheless, there is something honest about Einstein's attempt at popular writing: he does not gloss over difficulties. His theory was, to some

BIZARRE PREDICTIONS FROM GENERAL RELATIVITY, PREDICTIONS THAT EINSTEIN WAS WARY OF AT FIRST, HAVE STOLEN HIS THUNDER.

extent, all there in his book. The treatment just did not seem to work, and he knew it. He told a friend, the Swiss–Italian engineer Michele Besso, that it was "quite wooden". In later years, he joked with the Polish physicist Leopold Infeld that the description "generally understandable" on the book's cover should be changed to "generally not understandable".

Having given up on Einstein, I looked around and found much to choose from. As soon as Einstein had put his theory out, others took over and made it their own. Arthur Eddington, the UK astronomer who had measured the bending of light in a 1919 eclipse expedition, wrote a beautifully crafted mathematical treatise on the theory of space-time in 1923. Erwin Schrödinger, one of the fathers of quantum physics, came up with his own, more-conceptual rendition, *Space-Time Structure*, published in 1950. Fellow quantum pioneer Paul

Dirac's reticent personality comes across in his lecture notes on the theory, numbering just under 70 pages and published as *The General Theory of Relativity* in 1975. Euphoria and creativity pour out of the 1,200-page behemoth *Gravitation* (1973) by John Archibald Wheeler and his disciples Charles Misner and Kip Thorne. I was spoiled for choice.

Although I never used Einstein's book, it kept cropping up in my life. I have a penchant for second-hand bookshops and would keep finding translations, each with its own story. The French version was first translated by Jeanne Rouvière, a protégée of mathematician and politician Emile Borel, and subsequently expanded by Einstein's friend Maurice Solovine. The mathematician Tullio Levi-Civita, whose work had been instrumental in sucking Einstein into Riemannian geometry, recommended an engineer, Giuseppe Luigi Calisse, to do the Italian translation. The Russian version was translated by a Jewish logician-philosopher, Gregorius Itelson, who lived in Berlin and was beaten to death in 1926 by an anti-Semitic crowd.

Today, Einstein's book is a historical curio. I don't think anyone still reads it as he intended. There have been so many attempts at popularizing the theory, from practitioners and journalists, that anyone can find a book to their taste. And we have learnt much in the century since it was published: a popular book on relativity must now talk about the expanding Universe and the Big Bang, black holes and singularities. These bizarre predictions from general relativity, predictions that Einstein was wary of at first, have stolen his thunder.

Yet I can still see some fugitive magic in *Relativity*, despite its "wooden" tone. It conjures Einstein as the oracle presenting a theory to the world — one of the most revolutionary and profound theories of all time. ∎

Pedro Ferreira *is professor of astrophysics at the University of Oxford, UK, and author of* The Perfect Theory.
*e-mail: p.ferreira1@physics.ox.ac.uk*
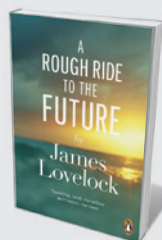
A Rough Ride to the Future
James Lovelock PENGUIN *2015*
Independent scientist James Lovelock gazes at Earth's past, present and future as the self-regulating system Gaia. Focusing on climate, he foresees humanity in 100 million years merged with artificial intelligence to survive a hotter Earth. (See Tim Lenton's review: *Nature* **508,** 41; 2014.)
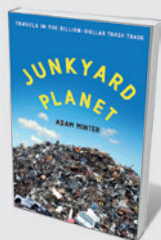
Junkyard Planet: Travels in the Billion-Dollar Trash Trade
Adam Minter BLOOMSBURY *2015*
One man's trash is another's treasure in Adam Minter's exploration of the US$500-billion global recycling trade. US waste tops the charts, and China's electronics-manufacturing industry sifts gold from mountains of e-waste. **Emily Banham**

Correspondence

Germline edits: heat does not help debate

Edward Lanphier and colleagues marshal a familiar and, we believe, ultimately counterproductive line of argument against genetically modifying the human germline (*Nature* **519,** 410–411; 2015, see also *Nature* http://doi.org/3xt; 2015).

In his essay *The Rhetoric of Reaction* (Belknap, 1991), the late economist Albert Hirschman describes three simplifying, stereotypical theses that are often invoked to preserve prevailing policies or practices. Lanphier *et al.* follow these same well-travelled tracks. They warn that a public outcry over the ethics of germline editing could hinder therapeutic applications of non-heritable genetic modification of somatic cells (the perversity thesis); that the technique might not deliver the anticipated benefits (the futility thesis); and that modifying genes in reproductive cells could be a slippery slope towards non-therapeutic genetic enhancement (the jeopardy thesis).

Such blinkered arguments can lead to sharply polarized views, obstructing the balanced and pragmatic societal discussion and careful studies that should precede possible applications of human germline modification (see D. Baltimore *et al. Science* **348,** 36–38; 2015).

Mika Martikainen, Ole Pedersen *Newcastle University, UK.*
mika.martikainen@newcastle.ac.uk

Germline edits: trust ethics review process

Edward Lanphier and colleagues contend that human germline editing is an unethical technology because it could have unpredictable effects on future generations. In our view, such misgivings do not justify their proposed moratorium (*Nature* **519,** 410–411; 2015).

When *in vitro* fertilization and preimplantation genetic diagnosis were first introduced, they had unpredictable consequences. Both went on to transform reproductive medicine.

Many nations already have ethics review processes that assess the risks of experiments on human embryos, with the prospect of even stricter evaluation standards as new fertility techniques come along (I. G. Cohen *et al. Science* **348,** 178–180; 2015).

There is no reason to close off whole avenues of controversial research when they have barely begun (see, for example, *Nature* http://doi.org/3xt; 2015). Germline editing is a revolutionary technology that potentially offers an enormous range of benefits to the next generation.

Julian Savulescu, Chris Gyngell, Tom Douglas *University of Oxford, UK.*
julian.savulescu@philosophy.ox.ac.uk

Lift sanctions now to save public health

The international accord on Iran's nuclear programme agreed this month in Switzerland offers no timeline for lifting international economic sanctions, which profoundly affect public health (see, for example, S. Shahabi *et al. Nature* **520,** 157; 2015). This must be urgently rectified because, as the Iranian health minister has stressed, it will take a year to restore public-health systems after sanctions are lifted.

By driving up prices and limiting the availability of medicines, sanctions are forcing people and clinics to use poor-quality, black-market medications. A shocking example still under investigation is last month's unexpected permanent loss of vision in 15 people after eye surgery in a Tehran clinic, with unknown numbers similarly affected at two other centres. The tragedy is thought to have arisen from a non-standard ampoule (see go.nature.com/oyudtj; in Persian).

Health care is a fundamental human right (see go.nature.com/xqtarv and go.nature.com/xuoeyb). Specialists in human rights, such as the United Nations special rapporteur on human rights in Iran, Ahmed Shaheed, should remind the US Congress and other Western governments of the importance of lifting sanctions and of a clear strategy to rapidly improve the country's public health.

Mehdi Aloosh *Montreal, Canada.*
Arash Aloosh *BI Norwegian Business School, Oslo, Norway.*
md_aloosh@hotmail.com

Double–blind peer review a double risk

Double-blind peer review of research papers is a worthy idea but has two flaws in practice (*Nature* **518,** 274; 2015).

First, most modern research builds on previous studies published by a limited number of groups. This makes it almost impossible to write a paper without revealing with near certainty who the authors are, even if they manage to avoid such giveaways as "we showed previously that … (No Longer Anonymous *et al.*, 2012)".

The second flaw is more serious. To function in our increasingly competitive research culture, in which misconduct is on the rise, researchers need to be aware of which labs can be trusted and which have a record of irreproducibility. If a highly regarded lab and one with a questionable reputation each submit reports of similar investigations, a good reviewer would be extra vigilant in assessing the less-reliable lab's study, even though the same evaluation standards would be upheld for both.

Double-blind peer review removes this crucial quality-control option, opening the way for mediocre and bad labs to clutter the literature with sub-standard science.

Thomas E. DeCoursey *Rush University Medical Center, Chicago, Illinois, USA.*
tdecours@rush.edu

Better policing for fishy catch data

Missing catch data are a big problem in achieving sustainability of world fisheries (*Nature* **519,** 280–282; 2015). The accuracy of catch data is also important — particularly for endangered fish species.

Take the common skate (*Dipturus batis*), now classified as critically endangered in the Red List of the International Union for Conservation of Nature. In 2009, the European Union put a total ban on fishing for this species and made it illegal to retain these fish on vessels, exchange them between boats, or land them.

We were therefore surprised to learn that official landing records of *D. batis* in 2014 by UK vessels in the United Kingdom and Europe, and by foreign vessels in UK ports, showed commercial catches totalling 1.8 tonnes. In 2011–13, the total recorded catch was even higher at 7.8 tonnes (all data by written request from the UK Marine Management Organisation).

The reliability of these figures is unclear, however, given that catch totals are compiled from assorted data collected at numerous UK ports. Local inaccuracies might be perpetrated through systematic errors in species identification, for example, or by mistakes in the codes allocated at market.

It is essential for the future of vulnerable species that catch data are properly checked by the authorities to monitor adherence to bans, and that punitive action is taken against contraventions.

David W. Sims, Samantha J. Simpson *Marine Biological Association of the United Kingdom, Plymouth, UK.*
dws@mba.ac.uk

NEWS & VIEWS

CANCER

An essential passenger with p53

Deletion of the *TP53* gene, an event seen in colorectal cancers, often occurs with co–deletion of a gene that encodes an enzyme subunit governing gene transcription. This creates a vulnerability ripe for therapeutic development. SEE LETTER P.697

JAMES E. BRADNER

There are increasing numbers of cancer drugs that successfully target cell-growth proteins that are activated by mutation. But when proteins that restrict cell growth, called tumour suppressors, are inactivated by mutation or deletion, few therapeutic strategies exist. Often, tumour-suppressor loss creates a pathway-specific cellular dependency, such as the reliance of BRCA-mutant breast-cancer cells on the enzyme poly(ADP-ribose) polymerase[1,2]. Regrettably, we have no general strategy for targeting tumour-suppressor loss, and nowhere is a creative solution to this challenge more urgently needed than in the context of the tumour suppressor p53. Alteration of the p53-encoding gene *TP53* is the most common genetic event observed among 12 of the most common cancer types (42% of tumours)[3]. Yet despite definitive biological characterization over decades of study, there is no drug that acts directly on the p53 pathway. On page 697 of this issue, Liu *et al.*[4] report that deletion of *TP53* in many colorectal cancers is often accompanied by deletion of one copy of a subunit of the enzyme RNA polymerase II. This co-deletion renders such hemizygous cells sensitive to polymerase inhibition by the natural toxin α-amanitin.

p53 is a transcription factor that activates the expression of cell-cycle checkpoint genes such as *CDKN1A* in response to DNA damage, thereby arresting cell proliferation. In cancer, alteration of *TP53*, either by mutation or deletion, promotes cell proliferation and survival. Mutant p53 lacks a capacity for transcriptional activation, and so direct-acting small molecules would need to have a compensatory, corrective function. Such molecules have proved elusive. Hemizygous deletion of *TP53* results in reduced abundance of p53, establishing a rationale for stabilizing remaining p53 by targeting the protein MDM2 — a ubiquitin-ligase enzyme that promotes p53 degradation[5]. Derivatives of the small molecule nutlin-3a, which disrupts the MDM2–p53 interaction, have recently transitioned to human clinical investigation. But there remains a pressing need for therapeutics that target p53 loss.

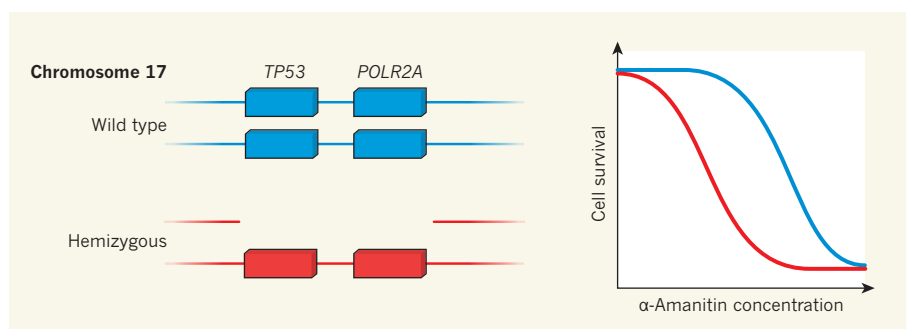Deletions in cancer genomes are not limited to tumour-suppressor genes, and often involve bystander or 'passenger' genes. There is an increasing appreciation that collateral genetic damage to passenger genes may create unique vulnerabilities in cancer cells. For example, a study[6] of the aggressive brain tumour glioblastoma multiforme found that the common causative mutation — deletion of both copies of the gene that encodes the metabolic enzyme enolase 1 (ENO1) — creates a heavy reliance on the remaining copies of the gene encoding the back-up enzyme enolase 2 (ENO2). Drugs that target ENO2 selectively killed ENO1-deleted cancer cells.

Another study[7] built on the observation that most human tumour cells harbour partial loss of at least 10% of their genome, typically as a result of the loss of whole chromosomes or chromosome arms. Among partially lost genes, the authors of that study found that hemizygous deletion of the gene *PSMC2*, which encodes a regulatory subunit of the essential proteasome protein-degradation complex, rendered ovarian-cancer cell lines highly sensitive to reduced PSMC2 expression (generated by the technique of RNA interference), both *in vitro* and *in vivo*.

Liu *et al.* extend this logic to study hemizygous loss of *TP53* in colorectal cancer. The authors find that regional deletions on chromosome 17 commonly involve both *TP53* and a neighbouring gene, *POLR2A*. This gene encodes the largest subunit (subunit A)

Figure 1 | Genetic vulnerability to transcriptional inhibition in colorectal cancer. Liu *et al.*[4] show that the large deletion on chromosome 17 that underlies many human colorectal cancers leads to hemizygous (one copy) loss not only of the *TP53* tumour-suppressor gene, but also of the gene *POLR2A*, which encodes a subunit of the essential enzyme RNA polymerase II. This event results in decreased abundance of the enzyme in the cancer cells, such that the cells are killed by much lower concentrations of the natural toxin α-amanitin than kill non-cancerous cells that express both copies of *POLR2A* (idealized survival curves are shown).

of RNA polymerase II, which transcribes messenger RNA from protein-coding genes, an essential function of all human cells. The authors report that hemizygous loss of *POLR2A* results in decreased abundance of RNA polymerase II in colorectal-cancer cells, leading them to suggest that these altered cancer cells could be more vulnerable to transcriptional inhibition than unaltered normal cells.

Using chemical and genetic approaches in defined colorectal-cancer cell lines, Liu *et al.* observed a crucial dependence on the remaining copy of *POLR2A* (Fig. 1). They found that hemizygosity for *POLR2A* renders colorectal cancer cells ten times more sensitive to α-amanitin, and highly sensitive to silencing of *POLR2A*-gene expression by RNA interference. Careful controls validated these findings, including desensitization to α-amanitin by restoration of normal *POLR2A* expression, and comparisons of engineered paired cell lines with or without hemizygous alteration of *POLR2A*.

To assess the potential clinical relevance of these findings, the authors conducted RNA-interference and α-amanitin-efficacy studies in immunocompromised mice bearing human tumours. To overcome the inadequacies of α-amanitin as a therapeutic agent (such as its systemic toxicity), an antibody–drug conjugate was used to deliver α-amanitin to tumours through binding to the cell-surface protein

EpCAM. In all studies, sensitivity of *POLR2A*-hemizygous cells to transcriptional inhibition was confirmed.

This discovery opens a therapeutic window for transcriptional inhibition in a common, genetically defined subtype of human cancer. Limitations of the approach include the likelihood of drug resistance arising through duplication (or further amplification) of the remaining intact *POLR2A* gene, and the reliance on the natural poison α-amanitin. Nevertheless, these studies establish powerful reagents and a strong rationale for studying inhibitors of transcription and transcriptional

signalling in cancer. Immediate next steps for this research are likely to include assessment of inhibitors of transcriptional kinase enzymes (for example, CDK7 inhibition by THZ-1) and modulators of chromatin-associated transcription factors (such as BET bromodomain inhibition by JQ1). Altogether, this study provides further validation that investigation of genomic loss can reveal previously unrecognized and actionable dependencies, establishing hope of a therapeutic inroad to managing *TP53* alteration in cancer. ∎

James E. Bradner *is in the Department*

of Medical Oncology, Dana-Farber Cancer Institute and the Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: james_bradner@dfci.harvard.edu

1. Bryant, H. E. *et al. Nature* **434**, 913–917 (2005).
2. Farmer, H. *et al. Nature* **434**, 917–921 (2005).
3. Kandoth, C. *et al. Nature* **502**, 333–339 (2013).
4. Liu, C. *et al. Nature* **520**, 697–701 (2015).
5. Wade, M., Li, Y.-C. & Wahl, G. M. *Nature Rev. Cancer* **13**, 83–96 (2013).
6. Muller, F. L. *et al. Nature* **488**, 337–342 (2012).
7. Nijhawan, D. *et al. Cell* **150**, 842–854 (2012).

IMMUNOLOGY

Stillbirth prevented by signal blockade

Bacterial infections during pregnancy can cause fetal death. The identification of a cell–signalling pathway induced during infection that recruits dangerous immune cells to the placenta may open up avenues to prevent this.

ANNA BAKARDJIEV

The birth of a dead baby during the second half of gestation is called stillbirth. It is an emotionally devastating experience for affected families and a major global health problem, especially in the developing world. Although there are many causes of stillbirth, maternal infections often play a key part, yet antimicrobial therapies are not effective. Writing in *The Journal of Clinical Investigation*, Chaturvedi *et al.*[1] report a cell-signalling pathway that leads to intrauterine accumulation of maternal immune cells during infection, which in turn triggers fetal death. The description of this mechanism may lead to the development of therapies to prevent stillbirth and other infection-related complications of pregnancy.

Infection during pregnancy differs from infection in non-pregnant individuals because mammalian gestation is a unique immunological situation. The fetus is genetically different from the mother, which brings the danger that the maternal immune system will mount a damaging immune response against this 'foreign' entity. Several cellular mechanisms prevent this[2]; for example, maternal immune cells that have the capacity to kill fetal cells, such as effector T cells (CD8[+] T cells), are kept away from the womb. Chaturvedi and colleagues show how one of these protective mechanisms is subverted when the mother is infected with *Listeria monocytogenes*. This bacterial pathogen causes a rare but often fatal food-borne disease (listeriosis) in humans and other mammals, and infection during pregnancy can lead to fetal death[3]. It has also been used for decades as a model system for studying intracellular pathogenesis and host defence[4].

Direct contact between fetal and maternal cells occurs almost exclusively in the placenta and the adjacent uterine implantation site, the decidua. The decidua is a fascinating tissue, in which fetal cells called trophoblasts are juxtaposed with maternal cells. It is of the utmost
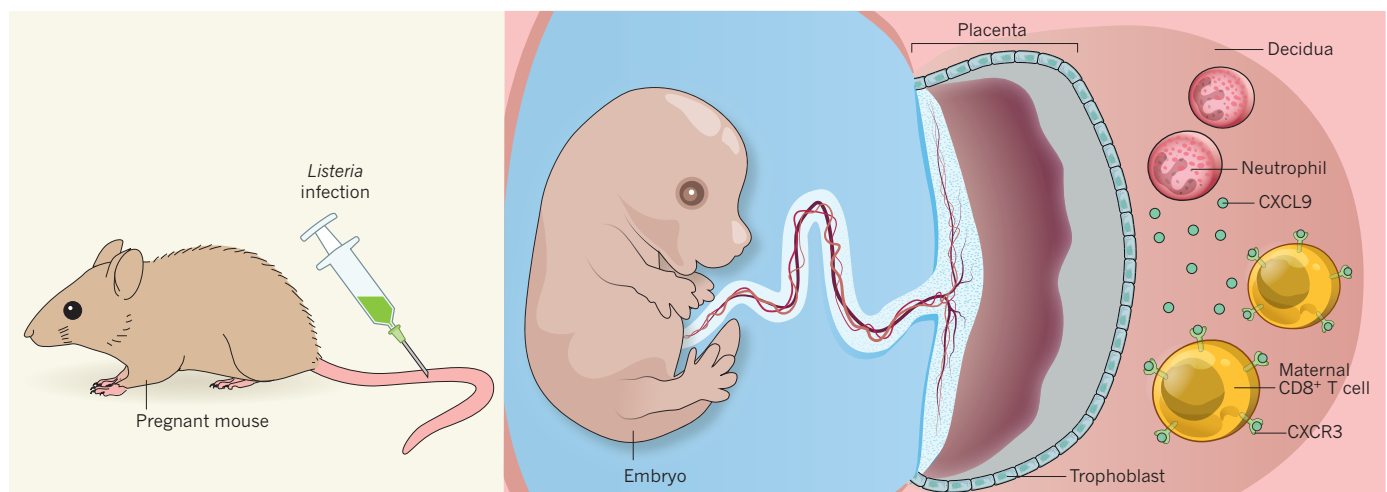

Figure 1 | Chemokine signalling triggers fetal death. During mammalian pregnancy, there is contact between the outermost fetal-derived cells of the placenta, the trophoblasts, and maternal cells in the uterine lining, which is called the decidua. Maternal immune cells that have the ability to recognize trophoblasts as being 'foreign' and to trigger an immune attack to kill the fetus are therefore typically kept out of the decidua. However, Chaturvedi *et al.*[1] show that intravenous infection of pregnant mice with the bacterium *Listeria monocytogenes* can lead to an influx of immune cells called neutrophils into the decidua; these express the chemokine molecule CXCL9, which attracts other immune cells, CD8[+] T cells, that express the CXCL9 receptor, CXCR3. In the authors' experimental model, in which these T cells all recognize fetal structures (antigens), this influx of cells leads to fetal death.

importance that maternal effector T cells that recognize fetal structures (antigens) be kept out of the decidua: if trophoblasts are killed by these cells, the life of the fetus is in danger. To help prevent this, the expression of genes encoding signalling molecules that attract T cells, such as CXCL9, are silenced in decidual tissue[5]. However, Chaturvedi and colleagues demonstrate that infection of pregnant mice with *L. monocytogenes* induces CXCL9 expression in the decidua, an influx of CD8[+] T cells, and fetal death (Fig. 1).

The authors used different strains of mice to pinpoint the connection between fetal-specific CD8[+] T cells and fetal demise. When they crossed female mice in which all CD8[+] T cells recognize a specific antigen (ovalbumin) with male mice in which all cells express that antigen (to provide a surrogate for fetal antigens and maternal CD8[+] T cells that recognize those antigens), they found that ovalbumin-specific T cells accumulated in the decidua after intravenous infection of pregnant animals with *L. monocytogenes*. These T cells express CXCR3, a receptor for CXCL9, and are recruited by CXCL9-secreting neutrophils (another class of immune cell) that move to the decidua after infection. Depleting neutrophils or blocking CXCR3 before or shortly after bacterial infection correlated with fewer ovalbumin-specific CD8[+] T cells in the decidua and no fetal death.

One of the caveats of this approach is that massive overexpression of a single antigen at the maternal–fetal interface does not reflect normal physiological conditions, nor does the presence of a large population of maternal T cells that all recognize that antigen with high affinity. This combination may lead to a more robust effect than occurs in normal pregnancies in which there are multiple antigens and a broad repertoire of T-cell specificities. Alternative techniques, such as using antigens derived from genetically different mice[6], might provide a more rigorous test of these ideas.

Is systemic listeriosis sufficient to trigger this signalling cascade, or is decidual infection required? Chaturvedi *et al.* used an *L. monocytogenes* mutant strain that is cleared more rapidly from maternal organs than wild-type bacteria[7], but that can grow in the murine placenta[8], suggesting that direct intrauterine invasion is necessary.

Further studies are needed to gather evidence for the role of this pathway in human pregnancy, because the anatomy of the maternal–fetal interface differs between mice and humans[9]. Nevertheless, Chaturvedi and colleagues' study is exciting, because it may have identified a mechanism that underlies the association of stillbirth with infection. Moreover, CXCR3 inhibitors are currently under investigation as therapeutic options for a broad range of diseases, including cancer, autoimmune conditions and infectious diseases[10]. Although more clinical trials are needed to optimize drug

safety and efficacy, especially where treatment of pregnant women would be considered, the authors' findings point to a new potential strategy for preventing stillbirth and other infection-related complications of pregnancy. ∎

Anna Bakardjiev *is at Benioff Children's Hospital, University of California, San Francisco, San Francisco, California 94143-0654, USA.*
e-mail: anna.bakardjiev@ucsf.edu

1. Chaturvedi, V. *et al. J. Clin. Invest.* **125,** 1713–1725 (2015).
2. Erlebacher, A. *Annu. Rev. Immunol.* **31,** 387–411 (2013).
3. Lamont, R. F. *et al. J. Perinat. Med.* **39,** 227–236 (2011).
4. Portnoy, D. A. *Mol. Biol. Cell* **23,** 1141–1145 (2012).
5. Nancy, P. *et al. Science* **336,** 1317–1321 (2012).
6. Wegorzewska, M. *et al. J. Immunol.* **192,** 1938–1945 (2014).
7. Auerbuch, V., Lenz, L. L. & Portnoy, D. A. *Infect. Immun.* **69,** 5953–5957 (2001).
8. Le Monnier, A. *et al. Infect. Immun.* **75,** 950–957 (2007).
9. Robbins, J. R. & Bakardjiev, A. I. *Curr. Opin. Microbiol.* **15,** 36–43 (2012).
10. Van Raemdonck, K., Van den Steen, P. E., Liekens, S., Van Damme, J. & Struyf, S. *Cytokine Growth Factor Rev.* http://dx.doi.org/10.1016/j.cytogfr.2014.11.009 (2014).

MICROBIOLOGY

Malaria runs rings round artemisinin

In parts of southeast Asia, malaria parasites are showing resistance to the active ingredient in artemisinin–based antimalarial drugs. Delineation of a cell–signalling pathway might help to explain this phenomenon. SEE LETTER P.683

JEREMY BURROWS

Imagine an unfinished jigsaw puzzle. There are two separate parts, which are clearly from the same picture, but the pieces linking the two fragments elude you. This is the situation facing the malaria research community. The first part of the puzzle is the fact that, in regions of southeast Asia, *Plasmodium falciparum* malaria parasites in the ring stage (an early blood-borne stage of the parasite's life cycle) are showing varying degrees of resistance to dihydroartemisinin[1,2], the active component of the artemisinin class of antimalarial drugs[3]. The second is the culprit in

this resistance, mutations in a protein called *P. falciparum* Kelch13 (PfKelch13), about which little is known[4], so that a mechanistic understanding of how resistance is conferred is missing. In this issue, Mbengue *et al.*[5] (page 683) provide some of the missing pieces to connect these two parts of the puzzle.

Drug resistance can arise in various ways: a biological target can change or mutate so as to reduce a drug's ability to bind to it; the drug can become physically separated from the target, such that it can no longer exert its effect; or the biological target can increase to high enough levels to overcome the presence of the drug. Mbengue and colleagues show
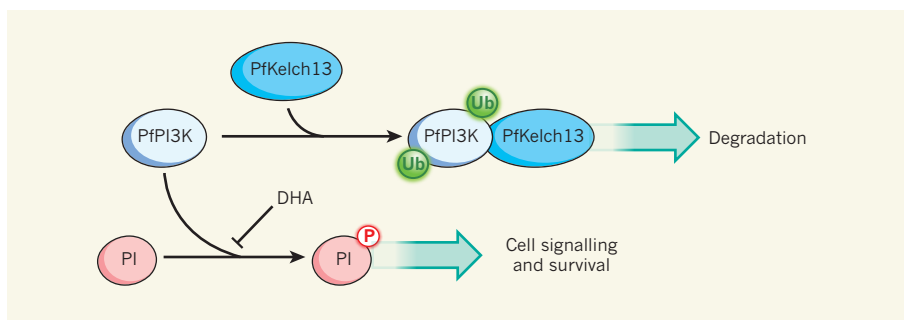

Figure 1 | Mechanism of resistance. Mbengue *et al.*[5] report that, in the malaria parasite *Plasmodium falciparum*, levels of the PfPI3K enzyme are regulated by PfKelch13, a protein that binds to PfPI3K and mediates the addition of ubiquitin groups (Ub), tagging PfPI3K for degradation. PfPI3K is required to phosphorylate (P) the phospholipid phosphatidylinositol (PI) to form phosphatidylinositol 3-phosphate, promoting cell signalling and survival. The authors show that dihydroartemisinin (DHA), the active ingredient of artemisinin-based antimalarial drugs, inhibits PfPI3K activity in ring-stage parasites. Mutations in PfKelch13 that prevent it from binding to PfPI3K can increase PfPI3K levels and so help parasites to overcome the effects of DHA.

that dihydroartemisinin (DHA) is a potent inhibitor of a *P. falciparum* enzyme called phosphatidylinositol 3-kinase (PfPI3K), which phosphorylates an important phospholipid, phosphatidylinositol (PI), to produce phosphatidylinositol 3-phosphate (PI3P) in ring-stage parasites[6]. However, mutations in PfPI3K do not correlate with resistance to DHA, indicating that this is not a main cause of resistance. The authors provide no evidence that the malaria parasite separates itself from DHA by destroying it or expelling it from its site of action, nor that the parasite hides in a protective state until DHA has dissipated[7]. This leaves the hypothesis that an increase in PfPI3K levels might confer resistance — but how is this linked to PfKelch13?

Clues about the function of PfKelch13 come from the roles of similar proteins in mammalian cells. The human equivalent of PfKelch13 binds to protein targets to promote molecular 'tagging', in which ubiquitin molecules are added to the substrate (a process called polyubiquitination). These ubiquitin-tagged protein complexes can trigger other biochemical processes, but they are also recognized and broken down by a cellular structure called the proteasome. Mbengue and co-workers reasoned that PfKelch13 might function in a similar manner in malarial ring-stage parasites, with PfPI3K as its target (Fig. 1). They hypothesized that mutations in PfKelch13 inhibit its association with PfPI3K. With less PfKelch13–PfPI3K binding, there would be less PfPI3K polyubiquitination, resulting in less degradation and higher cellular levels of PfPI3K. Indeed, the authors confirmed that increased levels of PfPI3K and PI3P were associated with certain PfKelch13 mutations. They had connected the two jigsaw pieces.

This work contains some surprises. First, it seems that this mechanism of action is specific to ring-stage parasites, implying that other reported biological consequences of DHA affect the parasite at other stages of its life cycle. By contrast, it is commonly assumed that a drug that kills parasites in different stages of its life cycle is driven by a single biological target, rather than different targets affecting different stages.

Second, the authors demonstrated that DHA inhibits PfPI3K more potently than it does related human kinases, and that this inhibition is reversible. DHA has a weak oxygen–oxygen (peroxide) bond, and the researchers provide data to suggest that this bond, along with other molecular features of DHA, is necessary for reversible PfPI3K inhibition. This is interesting, because all other mechanisms by which DHA kills parasites involve breaking the peroxide bond[8,9].

The third surprise is that DHA is different in structure from most types of molecule that inhibit kinases, although Mbengue *et al.* use modelling to provide some possible explanations for this discrepancy. It will be fascinating to carry out in-depth studies of interactions between DHA and PfPI3K.

Finally, PfKelch13 mutations correlate with only a slight increase in levels of PfPI3K in resistant parasites, but the extent of ring-stage resistance to DHA seems to be of a high magnitude, indicating that small alterations in PfPI3K levels can have a large effect on resistance. This suggests that resistance could be overcome by increasing the dose of artemisinin in drug combinations, to inhibit PfPI3K more potently — and such a study[10] has been independently conducted.

Many questions remain unanswered, the most important of which is why parasites survive when levels of PfPI3K and PI3P increase. Furthermore, exactly how DHA inhibits PfPI3K remains to be seen. How do PfPI3K and PfKelch13 bind together, and why do mutations hinder this binding? Finally, why does polyubiquitination increase following PfPI3K–PfKelch13 binding, and does tagging lead to intracellular signalling events or solely to destruction of the complex by the proteasome?

Will this knowledge of the ring-stage target of DHA help us to design better drugs, as the authors claim? Yes and no. Because resistance is due to factors other than PfPI3K mutation, new PfPI3K inhibitors will not only need to be very potent, but also, crucially, must show potency against the key ring-stage-resistant parasites. We know that mutations in PfKelch13 can lead to less PfKelch13–PfPI3K binding, and that this leads to resistance. Only time and a much greater understanding of the underlying biology will tell us whether the parasite might overcome PfPI3K inhibitors through mutations that further reduce binding between PfKelch13 and PfPI3K. From a drug-development perspective, given that there are other antimalarial drugs[11,12] in clinical development that have a weak peroxide bond but that differ from DHA, it will be intriguing to see whether they show inhibition of PfPI3K and resistance against the ring stage similar to those demonstrated by DHA[13].

It seems unlikely that this study will mark the completion of the malarial-resistance puzzle. A review of the field[2] has declared that "the debate continues". I am sure that it will. ∎

Jeremy Burrows *is at the Medicines for Malaria Venture, 1215 Meyrin, Switzerland.*
e-mail: burrowsj@mmv.org

1. Ding, X. C., Beck, H.-P. & Raso, G. *Trends Parasitol.* **27,** 73–81 (2011).
2. O'Neill, P. M., Barton, V. E. & Ward, S. A. *Molecules* **15,** 1705–1721 (2010).
3. Dondorp, A. M. *et al. N. Engl. J. Med.* **361,** 455–467 (2009).
4. Ariey, F. *et al. Nature* **505,** 50–55 (2014).
5. Mbengue, A. *et al. Nature* **520,** 683–687 (2015).
6. Vaid, A., Ranjan, R., Smythe, W. A., Hoppe, H. C. & Sharma, P. *Blood* **115,** 2500–2507 (2010).
7. Teuscher, F. *et al. J. Infect. Dis.* **202,** 1362–1368 (2010).
8. Meshnick, S. R., Taylor, T. E. & Kamchonwongpaisan, S. *Microbiol. Rev.* **60,** 301–315 (1996).

9. Haynes, R. K., Cheu, K.-W., N'Da, D., Coghi, P. & Monti, D. *Infect. Disord. Drug Targets* **13**, 217–277 (2013).
10. Das, D. *et al. Clin. Infect. Dis.* **56**, e48–e58 (2013).
11. Haynes, R. K. *et al. Angew. Chem. Int. Edn* **45**, 2082–2088 (2006).
12. Charman, S. A. *et al. Proc. Natl Acad. Sci. USA* **108**, 4400–4405 (2011).
13. Witkowski, B. *et al. Antimicrob. Agents Chemother.* **57**, 914–923 (2013).

This article was published online on 15 April 2015.

PALAEOCLIMATE

Northern push for the bipolar see–saw

Antarctic ice–core data show that abrupt changes of climate in the Northern Hemisphere in the last glacial period preceded associated shifts in Antarctica by about 200 years — indicating an oceanic coupling process. SEE LETTER P.661

TAS VAN OMMEN

Over the past 30 years, ice cores from Greenland, and subsequently from Antarctica, have progressively revealed a fascinating and unexpected picture of inter-hemispheric climate behaviour during the last glacial period, approximately 110,000 to 12,000 years ago. In the north, the glacial cold was punctuated by a series of abrupt warming events, each followed by cooling over several centuries before jumping back to cold, glacial conditions. These Dansgaard–Oeschger (DO) events, as they became known, have counterparts in Antarctic records, although with a different character: the Antarctic events show steady warming trends during Greenland's cold phases that peak and reverse to cooling trends when Greenland warms (Fig. 1).

As higher-resolution Antarctic ice-core records have been recovered, and tighter hemispheric synchronization of records has been achieved, it has become possible to unambiguously establish a one-to-one correspondence between DO events and the associated Antarctic temperature peaks. These peaks, observed in proxy form as variations of isotopic abundances of water in ice cores, are referred to as Antarctic Isotopic Maxima[1] (AIM). It has been difficult, however, to constrain the relative timing or phasing of the north–south events to better than a few centuries, so the presence of leads and lags, or of a directionality to the north–south coupling, has remained the subject of speculation — given the errors, the changes seemed synchronous. On page 661 of this issue, Buizert et al.[2] (WAIS Divide Project Members) report a clear Antarctic lag of around two centuries by using a new ice core from the West Antarctic Ice Sheet (WAIS) Divide. This establishes both a north-to-south

directionality and the relative timing of events, revealing a fairly slow connection that is consistent with hemispheric coupling by oceanic rather than atmospheric processes.

The underlying hypothesis for the observed hemispheric climate coupling, known as the bipolar see-saw[3,4], centres on the role of a large-scale ocean circulation in the Atlantic Ocean called the Atlantic meridional overturning circulation (AMOC). The AMOC carries heat from the Southern Hemisphere northwards until increasing salinity and cooling cause it to sink at high northern latitudes, forming North Atlantic Deep Water, which returns southwards at depth. Strengthening of the AMOC leads to warm temperatures in the North Atlantic and cool temperatures in the South Atlantic, with the converse occurring when the AMOC weakens.

This coupling might be expected to produce a simple inverse pattern of Antarctic and Greenland temperatures, but that is not the

case. The large thermal inertia of the Southern Ocean provides a heat reservoir in the south that accumulates the heat flow from the north. Including this effect in the bipolar-see-saw hypothesis gives a simple thermodynamic model that provides the basis for our understanding of the observations[5]. This model reveals that the abrupt DO warm events during the last glacial period reflect times when the AMOC intensified and was accompanied by heat removal from the south and cooling trends in Antarctica. Conversely, cold periods in Greenland and the north correspond to reduced AMOC heat transport from the south, accompanied by Antarctic warming trends.

This picture leaves questions of cause and effect unanswered, however. Potential drivers of AMOC variations include changes in freshwater balance, sea ice or ice shelves, any of which can initially affect deep-water formation in the north or the south. Buizert and colleagues suggest a straightforward interpretation of the north–south lag: the Northern Hemisphere AMOC shift triggers changes in the Southern Hemisphere, rather than the other way around. But, as others have noted[6,7] and the authors recognize, the concept of a 'trigger' may be poorly framed in a system of closely coupled oscillations — the system may be preconditioned for a state change by remote, as yet unidentified factors.

Buizert et al. find that the Antarctic temperature maxima occur, on average, 218 ± 92 years after DO warming transitions, and that the Antarctic minima lag behind the respective DO 'cold' transition by 208 ± 96 years.

The size of the time delays is interesting for two reasons. First, they are much longer than would be expected if atmospheric processes were responsible, pointing to an oceanic coupling mechanism. And second, the delays are the same whether the AMOC intensifies (the north warms) or reduces (the north cools). Both of these observations should provide new, substantial constraints on the processes responsible for transporting heat into the Southern Ocean and across the barrier of the Antarctic Circumpolar Current.

The findings from the WAIS Divide ice core approach the limits of such studies in determining the north–south delay. The errors are dominated by intrinsic uncertainties in working out the age of air trapped in the ice, and in detecting trend changes of ice-isotope abundances, which vary slowly and
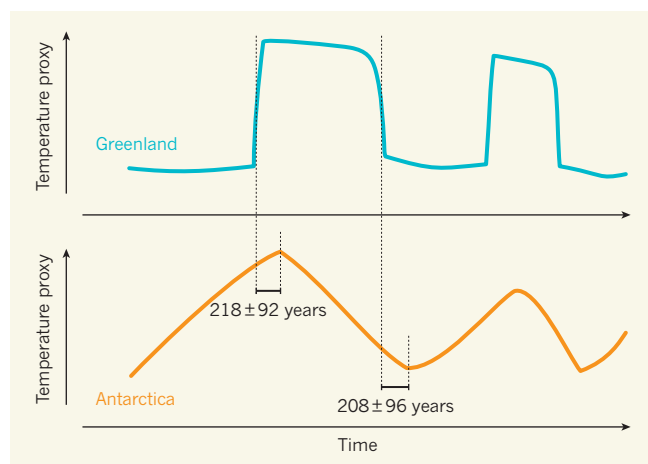

Figure 1 Paired warming events in Antarctica and Greenland. Antarctica warms when there are cool phases in Greenland, and cools when Greenland is in its warm state, as recorded by proxies based on isotopic abundances of water trapped in ice cores. Buizert et al.[2] find that Antarctic maxima occur on average 218 ± 92 years after abrupt warming in Greenland, whereas minima lag behind the cool state of Greenland by 208 ± 96 years. The graphs depict idealized trends, and are not based on recorded data.

suffer from climate-related noise. The authors averaged several AIM events to reduce noise, but although obtaining an average delay is informative, the question remains as to what extent individual DO–AIM pairs might vary in timing.

Studies of other Antarctic ice cores[8] reveal that the variation of temperature with time during AIM events has geographic variability, and that two phases are typically visible in the AIM profile during warming. They are accompanied by variations in other climate tracers that point to atmospheric-circulation changes that are not synchronous with Greenland events. Buizert *et al.* looked at atmospheric changes, and observed sea-salt variations during AIM that indicate synchronous changes in sea ice and temperature. Such variations may provide insight into changes in southern freshwater forcing and ocean feedbacks that affect the bipolar see-saw.

Establishing the relative timing between see-saw events in the two hemispheres is a big step forward, but the full extent of changes revealed by Antarctic ice cores, including the timing of changes in carbon dioxide level,

remains under-exploited. An integrated understanding of hemispheric climate coupling therefore awaits. Nevertheless, Buizert and colleagues' findings are particularly compelling in the light of recent indications[9] of a contemporary slowing of the AMOC, which has been anticipated. Predicting the global effects of such a change will pivot on our understanding of how the hemispheres communicate. ∎

Tas van Ommen *is in the Australian Antarctic Division at the Antarctic Climate and Ecosystems CRC, University of Tasmania, Hobart, Tasmania 7050, Australia.*
e-mail: tas.van.ommen@aad.gov.au

1. EPICA Community Members. *Nature* **444,** 195–198 (2006).
2. WAIS Divide Project Members. *Nature* **520,** 661–665 (2015).
3. Crowley, T. J. *Paleoceanography* **7,** 489–497 (1992).
4. Broecker, W. S. *Paleoceanography* **13,** 119–121 (1998).
5. Stocker, T. F. & Johnsen, S. J. *Paleoceanography* **18,** 1087 (2003).
6. Pedro, J. B. *et al. Clim. Past* **7,** 671–683 (2011).
7. Morgan, V. *et al. Science* **297,** 1862–1864 (2002).
8. Landais, A. *et al. Quat. Sci. Rev.* **114,** 18-32 (2015).
9. Rahmstorf, S. *et al. Nature Clim. Change* http://dx.doi.org/10.1038/nclimate2554 (2015).

MATERIALS SCIENCE

Semiconductors grown large and thin

Atomically thin layers of semiconductors called transition–metal dichalcogenides have been grown uniformly on the square–centimetre scale — paving the way for the ultimate miniaturization of electronic applications. SEE LETTER P.656

TOBIN J. MARKS & MARK C. HERSAM

The ubiquity of electronic devices today derives from the development of semiconductor wafers that have exceptional spatial uniformity. These wafers enable the production of highly integrated circuits, because each of the billions of constituent transistors behaves predictably, with differences between individual devices being among the smallest of any manufactured technology. Meanwhile, the miniaturization of transistors over the years has led researchers to consider the ultimate size limit: atomic-scale electronic devices. This limit has been reached in research laboratories with the fabrication of prototypes from atomically thin semiconducting materials[1]. But integrated circuits can be made from such devices only if atomically thin materials can be grown uniformly over large areas. On page 656 of this issue, Kang *et al.*[2] report a crucial step in this direction. They have achieved such uniformity on the wafer-scale — several square centimetres — for one of the

most promising classes of two-dimensional semiconductor.

Transition-metal dichalcogenides (TMDs) have the general formula MX_2, in which M is a metal such as molybdenum (Mo) or tungsten (W), and X can be sulfur, selenium or tellurium. They are semiconducting materials with 2D structures consisting of stacked, three-atom-thick X–M–X monolayers bound largely by interlayer van der Waals forces, roughly analogous to the structure of graphite. The appeal of TMDs has conventionally centred on their bulk forms[3], for applications as lubricants, energy-storage materials and catalysts. However, intense interest in the properties of atomically thin, 2D electronic materials such as graphene has extended to MX_2 materials, because they offer the possibility of high-performance and mechanically flexible transistors, light detectors, solar cells and light-emitting devices.

Most studies that have attempted to fabricate devices from atomically thin TMDs have used mechanically exfoliated forms (samples
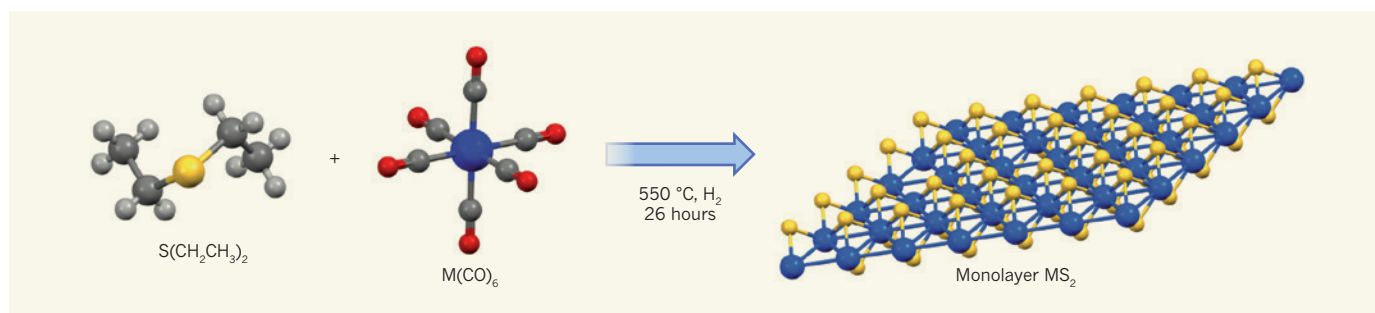
Figure 1 | Large-area growth of atomically thin, layered semiconductors. a, Kang et al.[2] report the fabrication of electronically uniform monolayers of semiconductors, MS_2 (M is molybdenum or tungsten), on the square-centimetre scale on silicon wafers coated with silica (SiO_2). The process involves exposing heated wafers to precursor compounds — $M(CO)_6$ and diethylsulfide ($S(CH_2CH_3)_2$) — at high temperatures in the presence of hydrogen gas. Sulfur, yellow; metal atoms, blue; oxygen, red; carbon, dark grey; hydrogen, light grey. Wafers are not shown.

prepared by stripping off a few atomic layers from the bulk material, for example by using self-adhesive tape)[4]. Alternatively, irregularly shaped monolayers have been used, or fragments a few layers thick on electrically insulating dielectric substrates[4]. The next step towards large-scale manufacturing will require practical pathways for coating wafer-scale areas with structurally and electronically uniform MX_2 monolayers on diverse substrates.

With this goal in mind, the methods of chemical vapour deposition (CVD) and metal–organic chemical vapour deposition (MOCVD) offer the ability to produce large-area coatings without the need for the expensive vacuum equipment required for other approaches. These techniques typically use volatile precursors to deliver the chemicals required to coat heated substrates, sometimes with extra energy provided in the form of a plasma or light. The practicality of CVD and MOCVD processes has led to their use on a vast scale for growing coatings on substrates as diverse as continuous glass sheets, drill bits and optoelectronic devices.

So far, MX_2 growth using these techniques has focused largely on thick, low-friction coatings, often with toxic and corrosive MF_6 precursors[5,6] (F is fluorine) and highly toxic hydrogen sulfide (H_2S) as a source of sulfur. Related approaches include aerosol-assisted MOCVD[7] and passing H_2S or sulfur vapour over hot metallic films or metal oxides (MO_3) pre-deposited on a substrate[8,9]. However, it seems unlikely that structurally and electronically uniform MX_2 monolayers can be reliably grown over large areas by any of the above approaches.

By contrast, Kang et al. describe an MOCVD process for growing MoS_2 and WS_2 monolayers that have excellent spatial uniformity on silicon wafers coated with silica (SiO_2). The process uses commercially available, volatile and relatively safe metal hexacarbonyl compounds, $Mo(CO)_6$ and $W(CO)_6$, as precursors of molybdenum and tungsten, respectively, and diethylsulfide ($S(CH_2CH_3)_2$) as a sulfur precursor (Fig. 1). The authors also added hydrogen gas to the precursor stream

to remove carbonaceous deposits that form during the growth process.

In this way, Kang and colleagues prepared MoS_2 monolayers across square-centimetre areas, and used them to make arrays of microscopic devices called field-effect transistors with a 99% device yield — only 2 out of 200 transistors characterized by the authors did not conduct. The electron mobility of the devices at room temperature was 30 square centimetres per volt per second — good for TMDs — and the mobility was minimally dependent on the dimensions of the transistors or their location on the wafer. The authors went on to demonstrate that silica can be deposited on the TMD monolayer after the first MOCVD growth step, and that the MOCVD process can then be repeated, resulting in multiple electrically isolated layers of atomically thin TMDs. Three-dimensional electronic architectures can thus be made, offering the enticing possibility of fabricating ultrahigh-density circuits — something that has not been possible using conventional silicon electronics.

Although the new findings constitute a substantial advance for atomically thin semiconductors, major issues must still be addressed before practical applications are possible. For example, the optimum growth conditions require a temperature of 550 °C to be maintained for 26 hours. This is too high a temperature to be used with currently available, flexible plastic substrates, and so a lower-temperature process is required, or a method for transferring large-area, atomically thin layers of TMDs from the growth substrate without introducing contamination, defects or wrinkles. The long growth time will probably also be an issue for high-throughput manufacturing.

Although the observed electron mobilities are respectable for TMDs, they are still more than ten times smaller than those of bulk crystalline silicon. Further research is also needed to find ways of 'doping' the TMDs — for example, by adding trace quantities of impurities — to control the type and concentration of charge carriers. Finally, the threshold voltage of the devices must be tuned to allow the development of low-power

electronic architectures, such as the widely used complementary metal-oxide–semiconductor (CMOS) technology.

The authors' 2D semiconductors open up opportunities for making devices and circuits that are far beyond conventional field-effect transistors. For example, because monolayer TMDs are atomically thin, they do not fully screen electric fields applied perpendicular to the monolayer; this property might allow the development of 'gate-tunable' heterojunction diodes for application in high-speed communications circuits[10]. Moreover, extended defects, such as grain boundaries in monolayer TMDs, can be manipulated using applied voltages. This might lead to the production of gate-tunable memristor devices, which are promising building blocks for 'non-volatile' computer memory and for neuromorphic (brain-like) circuit architectures[11]. Widespread access to uniform, large-area monolayer TMDs will help to accelerate progress in these emerging areas, allowing the full potential of 2D semiconductors to be quickly explored. ∎

Tobin J. Marks and **Mark C. Hersam** are in the Department of Materials Science and Engineering, the Department of Chemistry and the Materials Research Center, Northwestern University, Evanston, Illinois 60208, USA.
e-mails: t-marks@northwestern.edu; m-hersam@northwestern.edu

1. Jariwala, D., Sangwan, V. K., Lauhon, L. J., Marks, T. J. & Hersam, M. C. ACS Nano **8**, 1102–1120 (2014).
2. Kang, K. et al. Nature **520**, 656–660 (2015).
3. Chhowalla, M. et al. Nature Chem. **5**, 263–275 (2013).
4. Li, H., Wu, J., Yin, Z. & Zhang, H. Acc. Chem. Res. **47**, 1067–1075 (2014).
5. Scharf, T. W., Prasad, S. V., Mayer, T. M., Goeke, R. S. & Dugger M. T. J. Mater. Res. **19**, 3443–3446 (2004).
6. Doll, G. L., Mensah, B. A., Mohseni, H. & Scharf, T. W. J. Therm. Spray Tech. **19**, 510–516 (2010).
7. McCain, M. N., He, B., Sanati, J., Wang, Q. J. & Marks, T. J. Chem. Mater. **20**, 5438–5443 (2008).
8. Lee, Y.-H. et al. Adv. Mater. **24**, 2320–2325 (2012).
9. Song, I. et al. Angew. Chem. Int. Edn **53**, 1266–1269 (2014).
10. Jariwala, D. et al. Proc. Natl Acad. Sci. USA **110**, 18076–18080 (2013).
11. Sangwan, V. K. et al. Nature Nanotechnol. **10**, 10.1038/nnano.2015.56 (2015).

A multilevel multimodal circuit enhances action selection in *Drosophila*

Tomoko Ohyama[1]*, Casey M. Schneider-Mizell[1]*, Richard D. Fetter[1], Javier Valdes Aleman[1], Romain Franconville[1], Marta Rivera-Alba[1], Brett D. Mensh[1], Kristin M. Branson[1], Julie H. Simpson[1], James W. Truman[1], Albert Cardona[1]§ & Marta Zlatic[1]§

Natural events present multiple types of sensory cues, each detected by a specialized sensory modality. Combining information from several modalities is essential for the selection of appropriate actions. Key to understanding multimodal computations is determining the structural patterns of multimodal convergence and how these patterns contribute to behaviour. Modalities could converge early, late or at multiple levels in the sensory processing hierarchy. Here we show that combining mechanosensory and nociceptive cues synergistically enhances the selection of the fastest mode of escape locomotion in *Drosophila* larvae. In an electron microscopy volume that spans the entire insect nervous system, we reconstructed the multisensory circuit supporting the synergy, spanning multiple levels of the sensory processing hierarchy. The wiring diagram revealed a complex multilevel multimodal convergence architecture. Using behavioural and physiological studies, we identified functionally connected circuit nodes that trigger the fastest locomotor mode, and others that facilitate it, and we provide evidence that multiple levels of multimodal integration contribute to escape mode selection. We propose that the multilevel multimodal convergence architecture may be a general feature of multisensory circuits enabling complex input–output functions and selective tuning to ecologically relevant combinations of cues.

A single natural event, such as the attack of a predator, often presents multiple types of sensory cues, for example, visual, olfactory, tactile and noxious. These cues are detected separately by distinct sensors highly specialized for a single type of cue. Different sensory channels thus provide independent estimates of the same event that can be combined to improve sensitivity and reduce ambiguity[1–3]. Indeed, across the animal kingdom, multi-cue events often elicit enhanced reactions compared to single-cue stimuli[1–5].

The enhancement in reactions to multi-cue stimuli is a result of convergence and integration of information from multiple modalities[2,3,6]. Because neurons can compute non-linear functions of their inputs, where convergence happens, and how often, determines the repertoire of input–output functions[7]. In artificial neural networks, a single layer of nonlinear units has strict restrictions in possible responses to combinations of inputs[8,9]. Adding a second layer that processes combinations of outputs from the first layer enables the network to implement far more complex input–output functions[8–10] and thus to discriminate far better between different combinations of inputs. Similarly, in the nervous system, an architecture in which neurons that receive different combinations of inputs from two modalities in one level converge on neurons in downstream levels could enable better discrimination between different kinds of multimodal stimuli and selective tuning to specific ecologically relevant combinations of the two cues.

The structural and functional architecture of multisensory convergence across multiple levels of the sensory processing hierarchy is an open question for any large nervous system. Neurons that integrate information from multiple modalities have been most extensively studied in higher-order and motor-control areas[2,3,6,11–17]. Multisensory interactions have also been described at earlier stages of sensory processing[3,18–21]. However, it is unclear whether the multisensory neurons observed at distinct levels are structurally or functionally connected, and whether they contribute to the same behaviour.

The *Drosophila* larva is an excellent model system to study the structural and functional patterns of multimodal interactions across multiple levels of the sensory processing hierarchy and how these patterns contribute to behaviour. The relatively small size of its 10,000-neuron nervous system makes it amenable to circuit mapping using large scale electron microscopy reconstruction[22,23]. Powerful genetic tools enable functional studies of uniquely identifiable neurons[24–26].

Larvae have distinct sensory modalities with which they can gauge the degree of threat[27–31] and an array of distinct escape behaviours available for responding to threats, such as turning, digging, fast crawling and rolling[27,28,31–35]. The most vigorous escape sequence consists of two locomotion modes: rolling followed by fast crawling[28,32]. Rolling is the fastest—and presumably the most energetically costly[27,28]. A range of distinct aversive cues can evoke the fast crawling response without rolling, including mild nociceptive and strong mechanosensory stimulation[28,35]. In contrast, rolling is observed in extremely threatening situations, such as strong noxious stimulation[27,28,32,33,36], or predator attack[27,32], the latter of which provides multiple cues, including nociceptive and mechanosensory[27,31,32].

Here we show that mechanosensory combined with nociceptive cues synergistically increase the likelihood of rolling, and use this tractable model system to determine the anatomical and functional patterns of convergence supporting the synergy.

Multisensory cues enhance the selection of rolling

We developed an automated multisensory assay in which we activate larval nociceptive multidendritic class IV (MD IV) neurons thermogenetically[32,36] and larval mechanosensory chordotonal neurons using vibration[28,31] (Fig. 1a, b and Extended Data Fig. 1a–d).

[1]Howard Hughes Medical Institute Janelia Research Campus, 19700 Helix Drive, Ashburn, Virginia 20147, USA.
*These authors contributed equally to this work.
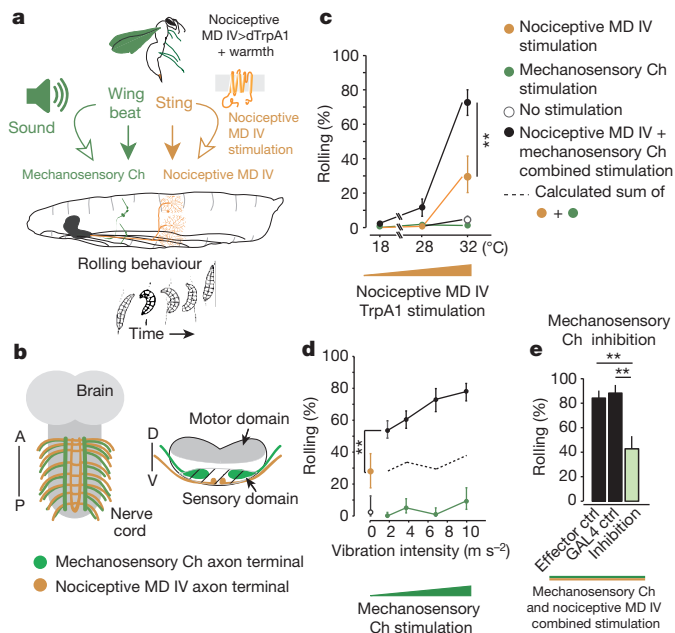§These authors jointly supervised this work.

Figure 1 | Mechanical and nociceptive stimulation synergistically trigger rolling. **a**, Schematic of a high-throughput multisensory assay for presenting stimuli similar to wasp attack. Speaker-emitted vibration (green) sensed by chordotonal (Ch) neurons replaces the mechanosensory stimulus. Thermogenetic activation of MD IV neurons (orange) using the transgenic temperature-sensitive channel, dTRPA1, and exposing *MD IV-GAL4>UAS-dTrpA1* larvae to warmth (30 °C or 32 °C), replaces the noxious sting. **b**, Chordotonal and MD IV neurons from each hemisegment project to distinct modality-specific ventrolateral and ventromedial regions of the nerve cord, respectively. A, anterior; P, posterior; D, dorsal; V, ventral. See Supplementary Table 1 for *n* (animals) in this and all other figures. See Supplementary Table 2 for a list of genotypes used in this and all other figures. **c**, **d**, Percentage of animals that rolled in response to mechanical (green, *MD IV-GAL4/+*), nociceptive (orange, *MD IV-GAL4>UAS-dTrpA1*), combined (black, *MD IV-GAL4>UAS-dTrpA1*), or no (white, *MD IV-GAL4/+*) stimulation, when either MD IV activation is increased, given constant vibration (1,000 Hz, 6.7 m s^{-2}; **c**), or vibration is increased, given constant nociceptor activation (32 °C; **d**). Error bars, 95% confidence interval; **$P < 0.0001$, χ^2 test with Holm–Bonferroni correction. **e**, Hyperpolarizing chordotonal neurons reduced rolling in response to bimodal mechanical and nociceptive stimulation (32 °C and 1,000 Hz, 6.7 m s^{-2} in this and subsequent figures) using the transgenic inward rectifying potassium channel Kir in *MD IV-LexA>lexAop-dTRPA1, ch-GAL4>UAS-Kir* animals. Ctrl, controls.

Different combinations of nociceptive and mechanosensory stimulation induced different likelihoods of the key escape sequences: rolling followed by fast crawling versus fast crawling alone. Nociceptor activation alone evoked a relatively low likelihood of rolling (Fig. 1c, d, Extended Data Fig. 1e and Supplementary Video 1) and a high likelihood of fast crawling[28] (Extended Data Fig. 1e, f). Vibration alone evoked only fast crawling[28,33] (Extended Data Fig. 1e, f) and essentially no rolling (Fig. 1c, d and Supplementary Video 2).

Combined with nociceptor activation, vibration increased the likelihood of rolling; the effect is dose-dependent and super-additive (synergistic) (Fig. 1c, d, Extended Data Fig. 1g and Supplementary Video 3). This vibration-induced facilitation of rolling was mediated through the mechanosensory chordotonal neurons (Fig. 1e).

First-order multisensory interneurons trigger rolling

We suspected that the information from the two modalities converges onto central neurons involved in the selection of rolling. To identify such neurons and thus determine where in the sensory processing hierarchy multisensory convergence occurs, we performed a behavioural screen for neurons whose thermogenetic activation triggers rolling (see Methods). We identified a 'hit', the *R72F11 Drosophila*

line, that drove GAL4 expression in neurons potentially early in the sensory processing hierarchy (Fig. 2a–c and Extended Data Fig. 2a, b). Activating the neurons in *R72F11* triggered rolling in a significant fraction of animals (Fig. 2a and Supplementary Videos 4 and 5), and inhibiting them (see Methods) significantly decreased rolling in response to bimodal stimulation (Fig. 2b).

R72F11 drives expression selectively in four lineage-related, segmentally repeated projection neurons with basin-shaped arbors in the ventral, sensory domain of the nerve cord; we therefore named them Basins-1–4 (Fig. 2c, d and Extended Data Fig. 2a–c). The dendrites of Basin-1 and Basin-3 span a ventrolateral domain of the nerve cord, where the mechanosensory chordotonal terminals are located[37,38] (Fig. 2c, d and Extended Data Fig. 2c, d). The dendrites of Basin-2 and Basin-4 span both the ventrolateral chordotonal domain and a ventromedial domain where the nociceptive MD IV terminals are located[37–39] (Fig. 2c, d and Extended Data Fig. 2c, d). We therefore asked whether the mechanosensory chordotonal and the nociceptive MD IV neurons directly converge on Basin-2 and Basin-4.

In an electron microscopy volume that spans 1.5 nerve cord segments, we reconstructed the chordotonal and MD IV arbors (Fig. 2e and Extended Data Fig. 2e, f) and all of their synaptic partners (a total of 301 neuronal arbors) (Fig. 2e, Extended Data Fig. 3a–d and Supplementary Table 4a–c). Amongst the reconstructed neurons we recognized the left and right Basin-1, -2, -3 and -4 homologues (Fig. 2d, f and Extended Data Fig. 2c, d, g, h).

Basin-1 and Basin-3 received many inputs (each >25 synapses and >15% of total input, on average) from chordotonal neurons, but very few (no more than 1% of total input synapses) from MD IV neurons (Fig. 2g and Extended Data Fig. 2i, j). Basin-2 and Basin-4 received many inputs from both chordotonal neurons (on average >20 synapses and >10% total input) and MD IV neurons (on average >20 synapses and >10% total input) (Fig. 2g and Extended Data Fig. 2i, j), each on distinct dendritic branches (Fig. 2h). Of all the 301 partners downstream of MD IV and chordotonal neurons, only Basin-2 and Basin-4 reproducibly received >5 synapses from both chordotonal and MD IV neurons (Fig. 2i), suggesting that they are probably key integrators of chordotonal and MD IV inputs.

To investigate whether the observed anatomical inputs from the sensory neurons onto Basins were functional and excitatory, we imaged calcium transients in response to MD IV or chordotonal activation (see Methods) collectively in all Basins (Extended Data Fig. 4a–d) or in individual Basin types (Fig. 3a, b), using lines that drive expression selectively in Basin-1 or Basin-4 neurons (Extended Data Fig. 4e–h).

In Basin-1, we observed calcium transients in response to vibration, but not in response to MD IV activation (Fig. 3a), consistent with the large number of synapses it receives from chordotonal neurons and the relatively few from MD IV neurons (Fig. 2g and Extended Data Fig. 2i). In Basin-4, we observed calcium transients in response to both vibration and MD IV activation (Fig. 3b), consistent with the large number of synapses it receives from both sensory types (Fig. 2g and Extended Data Fig. 2i). Basin-4 integrated the inputs from the two modalities, responding significantly more to bimodal than to unimodal stimulation (Fig. 3b, c and control in Extended Data Fig. 4i, j).

Next, we asked whether the multisensory Basin-4 interneurons contribute to rolling selection. Silencing Basin-4 neurons significantly decreased rolling in response to bimodal stimulation, indicating these neurons are involved in triggering rolling (Fig. 3d). Selective activation of the multisensory Basin-4 interneurons triggered rolling in a dose-dependent way, with strongest activation triggering rolling in 45% of animals (Fig. 3e).

We also wondered whether a second level of multimodal integration (that is, integration of information from distinct Basin types, that receive distinct combinations of chordotonal and MD IV inputs), enhances the selection of rolling. Indeed, co-activation of Basin-1 with the bimodal Basin-4 facilitated rolling (Fig. 3e), resulting in a
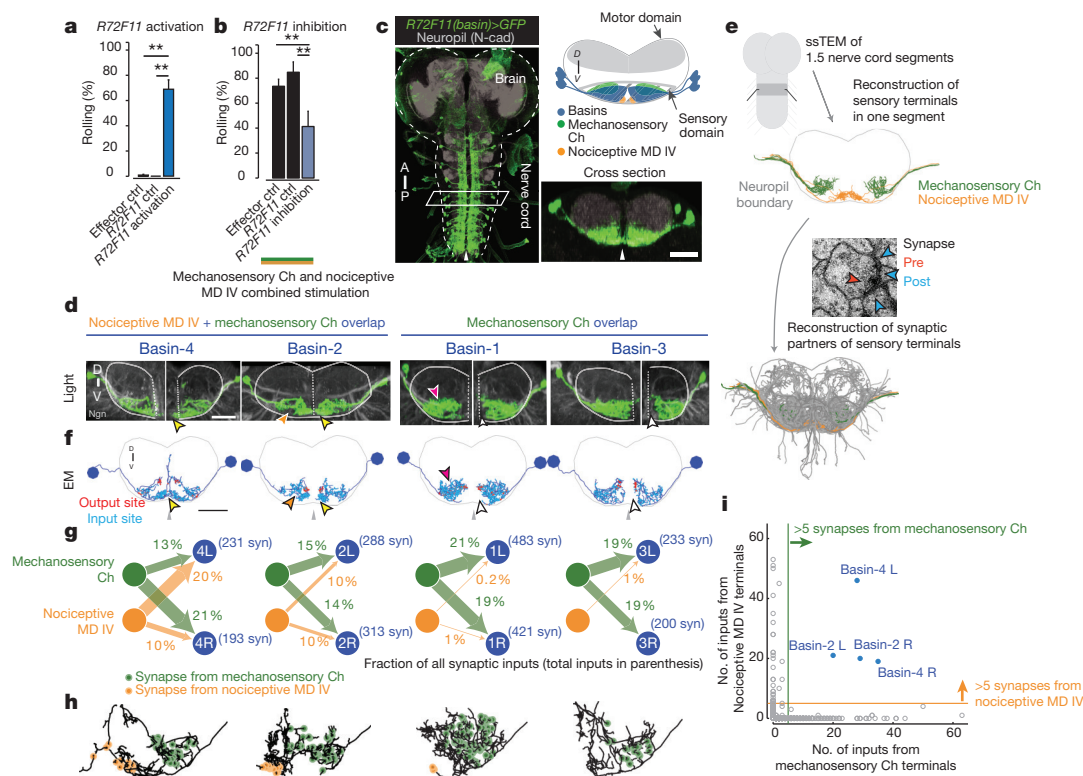
Figure 2 | Activation screen and electron microscopy reconstruction reveal first-order multisensory interneurons. a, Thermogenetic activation of neurons in *R72F11* (exposing *R72F11-GAL4>UAS-dTrpA1* larvae to 32 °C) triggered rolling. Error bars, 95% confidence interval; **$P < 0.0001$, χ^2 test with Holm–Bonferroni correction. b, Silencing neurons in *R72F11* significantly reduced rolling in response to bimodal stimulation (in *MD IV-LexA>lexAop-dTRPA1, R72F11-GAL4>UAS-Kir* animals). Ch, chordotonal. c, *R72F11* drives expression in Basin neurons (green, *R72F11-GAL4>UAS-GFP*). Scale bar, 25 μm. Arrowheads indicate midline. d, Four Basin types visualized with a FLP-based strategy are identifiable based on unique combinations of dendritic features: medial projection (yellow arrowhead); ventrolateral indentation (orange arrowhead); bushy dorsal arbor (red arrowhead); and absence of medial projection (white arrowhead). Solid white outlines, neuropil boundary;

dashed lines, midline. Neuropil is visualised with labelling against N-cadherin (N-cad) in c or neuroglian (Ngn) in d. Scale bar, 20 μm. e, Electron microscopy reconstruction of all direct synaptic targets of the chordotonal (eight left and eight right) and MD IV (three left and three right) neurons in the abdominal segment A3. ssTEM, serial section transmission electron microscopy. f, Amongst the reconstructed arbors, we recognized Basins, based on morphological features (arrowheads). Grey arrow, midline. Scale bar, 10 μm. g, The fraction of inputs onto Basins originating from chordotonal or MD IV terminals. h, Synaptic inputs from MD IV (orange circles) and chordotonal (green circles) segregate onto distinct dendritic branches of Basin-2 and Basin-4. i, Plot shows the number of synapses from chordotonal and MD IV terminals onto each of the 301 downstream partners. L, Basin homologue from the left hemisegment; R, Basin homologue from the right hemisegment.

significantly higher likelihood of rolling compared to activation of Basin-4 alone (70% versus 45%). Thus, information from distinct Basin types may converge again onto downstream neurons involved in triggering rolling.

Goro neurons in motor nerve cord trigger rolling

To identify potential sites of convergence of information from the different types of first-order Basin interneurons we looked at a 'hit' from the thermogenetic activation screen, *R69F06*, a *Drosophila* line that drove GAL4 expression in neurons that project far from the early sensory processing centres (Fig. 4a–c). Thermogenetic activation of neurons in *R69F06* triggered rolling in a high fraction of larvae (Fig. 4a and Supplementary Video 6), and inhibiting them significantly decreased rolling in response to bimodal stimulation (Fig. 4b).

R69F06 drives expression in a few neurons in the brain, in the suboesophageal zone (SEZ) and in a pair of thoracic neurons whose axons descend through the dorsal, motor domain of the nerve cord (Fig. 4c–e). Selectively activating the single pair of thoracic neurons (see Methods and Extended Data Fig. 5a–d) triggered rolling in 76% of larvae (Fig. 4e, f). We named these command-like neurons Goro (a romanization of the Japanese for rolling).

Activation of Basins evoked strong calcium transients in the Goro neurons, indicating that these cell types involved in the same behaviour are functionally connected (Fig. 4g, h and Extended Data Fig. 5e, f). To identify the shortest anatomical pathways from Basins to Goro that

might support the observed functional connectivity and to determine whether the information from distinct Basin types converges onto Goro, we again used electron microscopy reconstruction.

Nerve cord and brain pathways from Basins to Goro

We used a second electron microscopy volume (from a second larva) that spans the entire larval nervous system and therefore also includes Goro neurons (Fig. 5a and Supplementary Video 7). In the new volume, we reconstructed chordotonal, MD IV and Basin neurons from segment A1, as well as the Goro neurons (Fig. 5a–c, Extended Data Figs 2g–j and 3a–d, Supplementary Table 5a–c and Supplementary Video 8).

To find putative pathways from distinct Basin types to Goro neurons, we first reconstructed all neurons downstream of all axonal outputs from the four left and right Basin homologues from segment A1 (Fig. 5d, e, Supplementary Information (Supplementary Atlas section) and Supplementary Tables 5 and 6). We identified 31 pairs of reproducible downstream partners (Fig. 5d, e and Supplementary Information (Supplementary Atlas section)). Amongst these we identified second-order nerve cord interneurons that constitute the shortest pathways from Basins to Goro neurons (called A05q and A23g where 'A' stands for abdominal neuron; Fig. 5d). They receive inputs from distinct Basin types and synapse onto Goro neurons (Fig. 5d). Thus, information from distinct Basin types, that receive distinct combinations of MD IV and chordotonal inputs, converges onto Goro neurons—providing a second level of multimodal convergence.
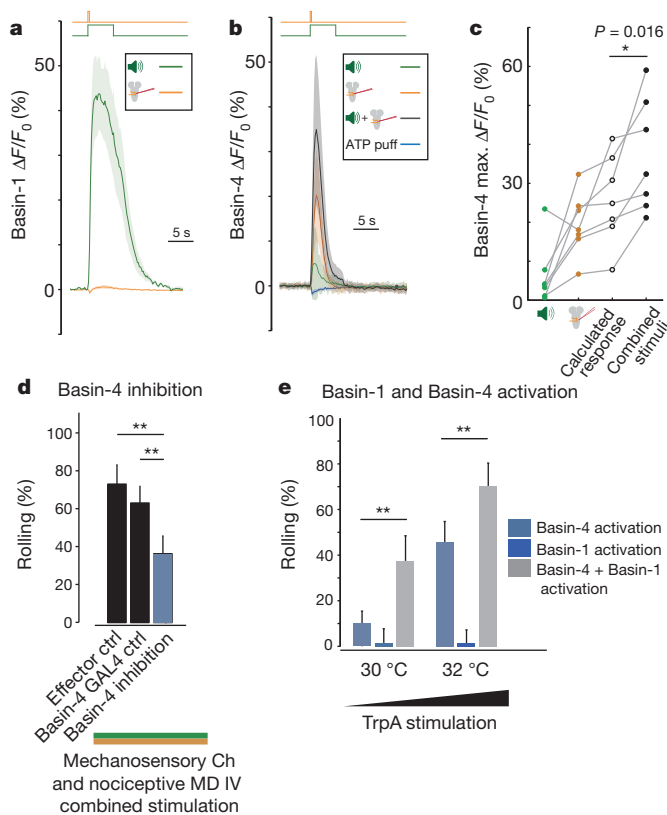
Figure 3 | Multiple levels of multimodal integration. a, Calcium transients (mean ± s.e.m.) are evoked in Basin-1 dendrites by vibration (1,000 Hz, 6.7 m s^{-2}, 5 s, symbolized by green speaker; green trace; $n = 7$ trials), but not by MD IV activation with transgenic ATP-gated channel, P2X$_2$, and local injection of ATP (symbolized by orange injection electrode; orange trace; $n = 6$ trials; *MD IV-lexA>laxAop-P2X$_2$; Basin-1-GAL4>UAS-GCaMP6*). **b,** Calcium transients (mean ± s.e.m.) are evoked in Basin-4 dendrites by vibration (green trace), nociception (orange trace) and combined stimulation (black trace; $n = 7$ triplet trials; *MD IV>P2X$_2$; Basin-4>GCaMP6*); but not by ATP injection in the absence of P2X$_2$ expression (blue trace, *MD IV>+; Basin-4>GCaMP6*). **c,** Peak $\Delta F/F_0$ from **e** shows Basin-4 response to combined stimulation (filled black circles) is significantly higher (paired t-test) than the calculated sum of unisensory responses (clear black circles). **d,** Silencing Basin-4 neurons significantly reduced rolling evoked by bimodal stimulation (in *MD IV-LexA>lexAop-dTRPA1, Basin-4-GAL4>UAS-Kir* animals) compared to controls. Ch, chordotonal. **e,** Thermogenetic activation of Basin-4 neurons evoked rolling (exposing *Basin-4-GAL4>UAS-dTrpA1* animals to 30 °C or to 32 °C). Basin-1 facilitated rolling in combination with Basin-4 (exposing *Basin-4-GAL4, Basin-1-GAL4>UAS-dTrpA1* animals to 30 °C or to 32 °C). Error bars, 95% confidence interval; **$P < 0.0001$, χ^2 test with Holm–Bonferroni correction.

Figure 4 | Basins are functionally connected to command-like Goro neurons. a, Thermogenetic activation of neurons in *R69F06* triggered rolling (exposing *R69F06-GAL4>UAS-dTrpA1* to 32 °C). **b,** Silencing neurons in *R69F06* significantly reduced rolling in response to bimodal stimulation (*MD IV-LexA>lexAop-dTRPA1, R69F06-GAL4>UAS-Kir*). Ch, chordotonal. **c–e,** Images of neurons (green) in *R69F06>GFP* (**c**), *R69F06,tsh-Gal80>GFP* (**d**) and *Goro>dTRPA1-myc* (*R16F11-LexA>lexAop-FLP, R69F06-GAL4>UAS-FRT-STOP-FRT-dTRPA1-myc*) (**e**) animals. Scale bar, 25 μm. Arrowheads indicate midline. **f,** Thermogenetic activation of the single pair of Goro neurons triggered rolling in 76% of animals (exposing *Goro>dTRPA1-myc* from the *R16F11-LexA>lexAop-FLP, R69F06-GAL4>UAS-FRT-STOP-FRT-dTRPA-myc* cross to 32 °C); thermogenetic activation of only the brain and SEZ neurons in *R69F06* did not (exposing *R69F06, tsh-GAL80>dTRPA1* to 32 °C). Error bars, 95% confidence interval; **$P < 0.0001$, χ^2 test with Holm–Bonferroni correction. **g,** Basin activation by local ATP injection (in *72F11>P2X$_2$, R69F06>GCaMP* animals) evoked calcium transients (mean ± s.e.m.) in Goro dendrites (red trace, $n = 9$ trials); ATP injection in the absence of P2X$_2$ (in *72F11>+; R69F06>GCaMP* animals) did not (grey trace, $n = 8$ trials). **h,** Summary of the identified functional pathway involved in triggering rolling.

Ten distinct second-order projection neuron types downstream of Basins ascend to the brain (Fig. 5e and Supplementary Information (Supplementary Atlas section)). Some of these integrate Basin information across multiple distal segments of the body, either exclusively from a single Basin type (for example, A00c-a6, Fig. 5e), or from distinct Basin types (that receive distinct combinations of sensory inputs; for example, A00c-a4 and A00c-a5, Fig. 5e). Then, distinct second-order PNs, that receive distinct combinations of Basin inputs (and therefore distinct combinations of mechanosensory and nociceptive inputs), re-converge again on third-order interneurons in the brain (Fig. 5f). Thus, following convergence of local mechanosensory and nociceptive information from a single segment onto multisensory Basins, global mechanosensory and multisensory information from multiple segments is integrated within the brain pathway.

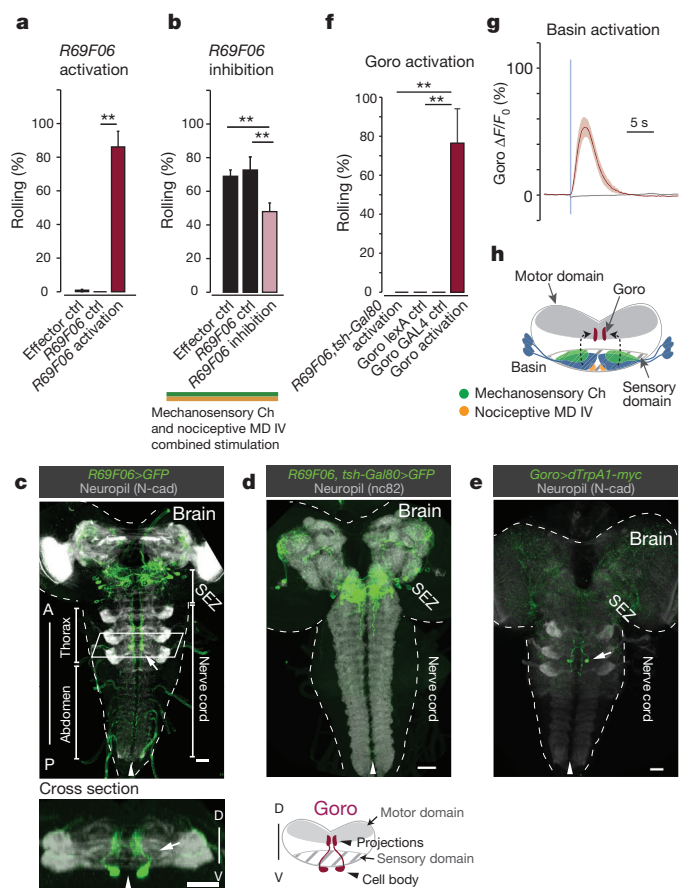By tracing upstream of Goro dendritic inputs (Fig. 5f and Supplementary Atlas) we identified brain neurons which send descending axons that synapse onto Goro neurons. Tracing downstream of a multisensory second-order ascending projection neuron (A00c-a4), we identified third-order projection neurons connecting the ascending pathways from Basins to a descending path onto Goro neurons (Fig. 5f). Thus, the activity of the command-like Goro neurons may be modulated by the more local multisensory and unisensory information via the nerve cord Basin–Goro pathway and by the global body-wide nociceptive and mechanosensory multisensory information via the brain Basin–Goro pathways (Fig. 5d–f).

We also identified a third-order SEZ feedback neuron that receives convergent body-wide mechanosensory and multisensory information and descends through the nerve cord sensory domain (Fig. 5g). The SEZ feedback neuron synapses onto the first-order (Basins) and second-order neurons from both the nerve cord (A05q) and brain (A00c) Basin–Goro pathways (Fig. 5g). Both the nerve cord and brain
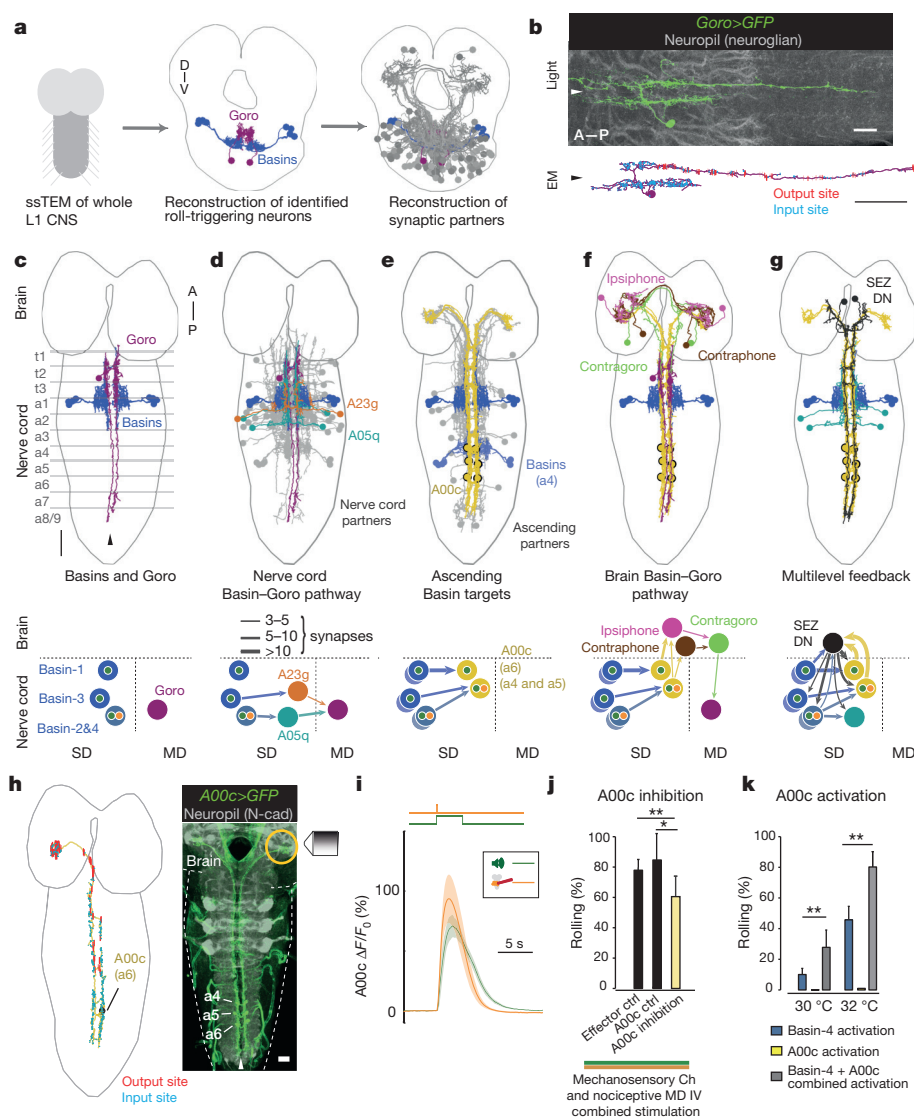
Figure 5 | Electron microscopy reconstruction reveals multiple levels of multimodal convergence within Basins-to-Goro pathways. **a**, Electron microscopy reconstruction of the Basin–Goro anatomical pathways. ssTEM, serial section transmission electron microscopy. **b**, Confocal image (green) and electron microscopy (EM) reconstruction of a single Goro neuron (magenta; see also Supplementary Video 8). Arrowheads indicate midline. Scale bars, 20 μm. **c–h**, Reconstructed interneurons (top) and their connectivity (bottom). MD, motor domain; SD, sensory domain. **c**, Reconstructed Basins and Goro neurons. **d**, Mechanosensory Basin target A23g and multisensory Basin target A05q converge onto Goro neurons. Grey, all other Basin targets that project to the nerve cord. **e**, Distinct A00c (yellow) ascending projection neurons integrate information from distinct combinations of Basins across multiple segments. Grey, all other Basin targets that ascend to the brain. **f**, Third-order brain projection neurons, ipsiphone and contraphone, receive inputs from the A00c neurons and synapse onto brain neurons (contragoro) that send descending axons into the nerve cord and synapse onto Goro. **g**, Feedback neurons from SEZ. DN, descending neuron. **h**, Electron microscopy reconstruction of a single A00c-a6 neuron and confocal image of all three pairs of A00c-a4/a5/a6 neurons in line *R71A10*. **i**, Calcium transients (mean ± s.e.m.) in the terminals of all A00c cells (in *MD IV>P2X₂, R71A10>GCaMP* animals) shows they respond to MD IV activation by local ATP injection (orange trace, $n = 12$ trials) and vibration (green trace, $n = 4$ trials). **j**, Silencing A00c neurons significantly reduces rolling in response to bimodal stimulation (in *MD IV-LexA>lexAop-dTRPA1, R71A10-GAL4>UAS-Kir* animals). Ch, chordotonal. **k**, A00c neurons facilitate rolling in combination with Basin-4 (exposing *Basin-4-GAL4, 71A10-GAL4>UAS-dTrpA1* animals to 30 °C or to 32 °C). Error bars, 95% confidence interval; **$P < 0.0001$, χ^2 test with Holm–Bonferroni correction.

Basin–Goro pathways may therefore be jointly regulated based on integrated global multisensory information.

Function of distinct pathways in triggering rolling

Next we wanted to explore the functional role of the nerve cord and brain Basin–Goro pathways. Basin activation could activate Goro neurons in the absence of the brain, suggesting the nerve cord Basin–Goro pathway is excitatory and sufficient for activating Goro neurons and triggering rolling (Extended Data Fig. 6a, b). Consistent with this idea, Basin activation evoked calcium transients in their nerve-cord targets, the A05q neurons, and A05q activation evoked calcium transients in Goro neurons, (Extended Data Fig. 6c–g). Furthermore, thermogenetic activation of the neurons in a line that drives expression, amongst others, in the A05q neurons triggered rolling (Extended Data Fig. 6c, h).

Calcium imaging in the terminals of three Basin-target neurons that ascend to the brain (A00c-a6, A00c-a5 and A00c-a4) (Fig. 5h) revealed that, collectively, they respond to vibration, to MD IV activation, and to Basin activation, suggesting that this connection is also excitatory (Fig. 5i and Extended Data Fig. 6i, j).

Silencing the A00c neurons decreased rolling in response to bimodal stimulation (Fig. 5j), and their co-activation with Basin-4 facilitated rolling (Fig. 5k). Therefore, downstream from early local multisensory integration by Basin-4, additional levels of integration of global mechanosensory and multisensory information appear to further facilitate the transition to rolling behaviour.

Differential effects on triggering roll and crawl

The rolling response triggered by multisensory cues (or by strong nociceptive cues alone) is followed by fast crawling[28] (Extended Data Fig. 1e, f). Similarly, optogenetic[40] activation of the first-order multisensory Basin-4 neurons triggered both locomotor modes; rolling followed by fast crawling (Extended Data Fig. 7a–c). However, optogenetic activation of the Goro neurons triggered only rolling, but not fast crawling, suggesting that they act as dedicated command-like neurons for rolling (Extended Data Fig. 7a–c). This also suggests that the act of rolling itself is insufficient to trigger fast crawling. In the future it will be interesting to determine how all of the 31 novel neuron types directly downstream of the Basins identified in our electron microscopy reconstruction (Supplementary Information (Supplementary Atlas section)) contribute to the selection of the two locomotor modes, rolling and crawling, in a defined sequence.

Discussion

By combining behavioural and physiological studies with large-scale electron microscopy reconstruction we mapped a multisensory circuit
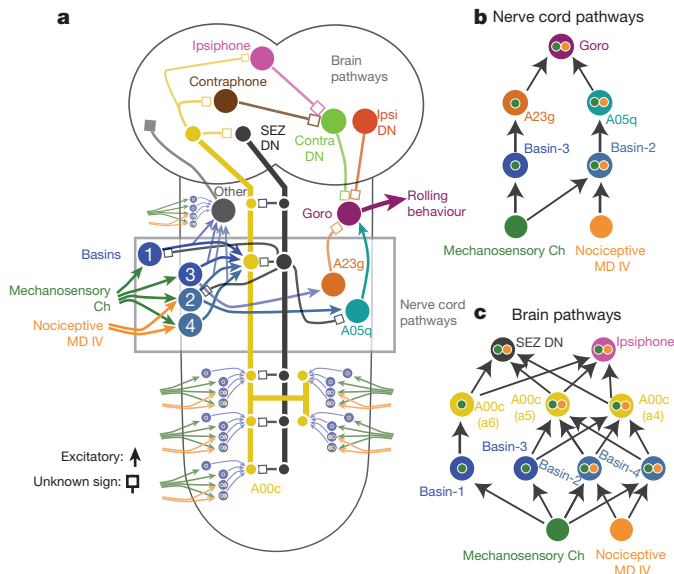
Figure 6 | Summary of the multilevel multimodal convergence circuit for rolling. a, Whole-animal circuit. Ch, chordotonal. **b, c**, Specific nerve cord (**b**) and brain (**c**) pathways Arrow, excitatory connections. Square, connections of unknown sign.

that mediates the selection of the fastest mode of escape locomotion (rolling) in *Drosophila* larva (Fig. 6a, b). We find that mechanosensory and nociceptive sensory neurons converge on specific types of first-order multisensory interneurons that integrate their inputs. Then, interneurons that receive distinct combinations of mechanosensory and nociceptive inputs converge again at multiple levels downstream, all the way to command-like neurons in the nerve cord. Activating just a single type of first-order multisensory interneuron triggers rolling probabilistically. Co-activation of first-order interneurons that receive distinct combinations of mechanosensory and nociceptive inputs increases rolling probability. Thus, action selection starts at the first-order multisensory interneurons and multiple stages of multimodal integration in the distributed network enhance this selection.

Given that spurious firing from distinct sensors is uncorrelated, whereas event-derived signals will be temporally correlated across the sensory channels, multimodal integration even at a single level improves the signal-to-noise ratio[1,3,41]. Multilevel multimodal integration can offer additional advantages. Theoretical studies show that a multilevel convergence architecture enables more complex input–output relationships[7–10]. Similarly, the multilevel multimodal convergence architecture described here could offer better discrimination between different kinds of multisensory events. The weights in such networks could be tuned either through experience or through evolution to respond selectively to highly specific combinations of two cues. Using a simple model, we can demonstrate (Extended Data Fig. 8a–k) that compared to early-convergence a multilevel architecture could specifically enhance the selection of the fastest escape mode in the most threatening situations, either in response to weak multimodal or strong unimodal nociceptive cues.

The multilevel multimodal convergence architecture may be a general feature in multisensory integration circuits, enabling complex response profiles tunable to specific ecological needs. For example, physiological studies in mammals have identified multisensory neurons that integrate the same cues at several stages in the sensory processing hierarchy[18,19,21,42,43], although it is unclear whether the multisensory neurons at distinct levels are causally related to the same behaviour. Due to the size of networks involved, synaptic-level resolution studies of the underlying convergence architecture across multiple levels were unattainable.

In addition to the multilevel multimodal feed-forward convergence motif, our electron microscopy reconstruction revealed higher-order and local feedback neurons (Figs 5g and 6a and Supplementary Information (Supplementary Atlas section)). Recent theoretical models of multisensory integration suggest that the output of individual multisensory neurons is normalized by the activity of other multisensory neurons in that population, but the anatomical implementation of such feedback has not been identified[2,44]. Some of the feedback neurons in the multisensory circuit described in this study may have roles in such normalization computations.

Another circuit motif revealed by our study is the divergence of sensory information into nerve cord and ascending brain pathways and subsequent re-convergence of the shorter and the longer pathway onto the same command-like neurons in motor nerve cord (Goro). The nerve cord pathway integrates nociceptive and mechanosensory information from a local region of the body (few segments), whereas the ascending brain pathway integrates the information across all body segments and provides a means of modulating command-like neuron activity based on global body-wide nociceptive and mechanosensory information. The multisensory circuit described here in a genetically tractable model system provides a resource for investigating in detail how multiple brain and nerve cord pathways interact with each other and contribute to the selection of different modes of locomotion (rolling and crawling) in a defined sequence.

The electron microscopy volume spanning the entire insect nervous system acquired for this study can be used to map circuits that mediate many different behaviours. Combining information from a complete wiring diagram with functional studies has been very fruitful in the 302-neuron nervous system of *C. elegans*[45,46]. Recently, similar approaches have been applied to microcircuits in smaller regions of larger nervous systems[47–50]. In this manuscript, we demonstrate that relating local and global structure to function in a complete nervous system is now possible for the larger and more elaborate nervous system of an insect.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Stein, E. B. & Meredith, M. A. *The Merging of the Senses* (MIT press, 1993).
2. Fetsch, C. R., DeAngelis, G. C. & Angelaki, D. E. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Rev. Neurosci.* **14,** 429–442 (2013).
3. van Atteveldt, N., Murray, M. M., Thut, G. & Schroeder, C. E. Multisensory integration: flexible use of general operations. *Neuron* **81,** 1240–1253 (2014).
4. Stein, B. E., Meredith, M. A., Huneycutt, W. S. & McDade, L. Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J. Cogn. Neurosci.* **1,** 12–24 (1989).
5. McMeniman, C. J., Corfas, R. A., Matthews, B. J., Ritchie, S. A. & Vosshall, L. B. Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell* **156,** 1060–1071 (2014).
6. Stein, B. E. & Stanford, T. R. Multisensory integration: current issues from the perspective of the single neuron. *Nature Rev. Neurosci.* **9,** 255–266 (2008).
7. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **2,** 303–314 (1989).
8. Duda, R. O. E., H. P& Stork, D. G. *Pattern Classification* (John Wiley and Sons, 2001).
9. Hastie, T., Friedman, J. & Tibshirani, R. *The Elements of Statistical Learning* Vol. 2 (Springer, 2009).
10. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497,** 585–590 (2013).
11. Fetsch, C. R., Pouget, A., DeAngelis, G. C. & Angelaki, D. E. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neurosci.* **15,** 146–154 (2012).
12. Meredith, M. A. & Stein, B. E. Interactions among converging sensory inputs in the superior colliculus. *Science* **221,** 389–391 (1983).
13. Meredith, M. A. & Stein, B. E. Descending efferents from the superior colliculus relay integrated multisensory information. *Science* **227,** 657–659 (1985).
14. Olcese, U., Iurilli, G. & Medini, P. Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron* **79,** 579–593 (2013).
15. Homberg, U., Christensen, T. A. & Hildebrand, J. G. Structure and function of the deutocerebrum in insects. *Annu. Rev. Entomol.* **34,** 477–501 (1989).

16. Milde, J. J. & Strausfeld, N. J. Cluster organization and response characteristics of the giant fiber pathway of the blowfly *Calliphora erythrocephala*. *J. Comp. Neurol.* **294,** 59–75 (1990).

17. Mu, L., Bacon, J. P., Ito, K. & Strausfeld, N. J. Responses of *Drosophila* giant descending neurons to visual and mechanical stimuli. *J. Exp. Biol.* **217,** 2121–2129 (2014).

18. Basura, G. J., Koehler, S. D. & Shore, S. E. Multi-sensory integration in brainstem and auditory cortex. *Brain Res.* **1485,** 95–107 (2012).

19. Jain, R. & Shore, S. External inferior colliculus integrates trigeminal and acoustic information: unit responses to trigeminal nucleus and acoustic stimulation in the guinea pig. *Neurosci. Lett.* **395,** 71–75 (2006).

20. Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A. & Schroeder, C. E. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* **53,** 279–292 (2007).

21. Shore, S. E. Multisensory integration in the dorsal cochlear nucleus: unit responses to acoustic and trigeminal ganglion stimulation. *Eur. J. Neurosci.* **21,** 3334–3348 (2005).

22. Cardona, A. *et al.* An integrated micro- and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. *PLoS Biol.* **8,** e1000502 (2010).

23. Saalfeld, S., Fetter, R., Cardona, A. & Tomancak, P. Elastic volume reconstruction from series of ultra-thin microscopy sections. *Nature Methods* **9,** 717–720 (2012).

24. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in Drosophila. *Proc. Natl Acad. Sci. USA* **105,** 9715–9720 (2008).

25. Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in Drosophila. *Genetics* **186,** 735–755 (2010).

26. Li, H. H. *et al.* A GAL4 driver resource for developmental and behavioral studies on the larval CNS of *Drosophila. Cell Rep.* **8,** 897–908 (2014).

27. Hwang, R. Y. *et al.* Nociceptive neurons protect *Drosophila* larvae from parasitoid wasps. *Curr. Biol.* **17,** 2105–2116 (2007).

28. Ohyama, T. *et al.* High-throughput analysis of stimulus-evoked behaviors in Drosophila larva reveals multiple modality-specific escape strategies. *PLoS ONE* **8,** e71706 (2013).

29. Tsubouchi, A., Caldwell, J. C. & Tracey, W. D. Dendritic filopodia, Ripped Pocket, NOMPC, and NMDARs contribute to the sense of touch in *Drosophila* larvae. *Curr. Biol.* **22,** 2124–2134 (2012).

30. Yan, Z. *et al. Drosophila* NOMPC is a mechanotransduction channel subunit for gentle-touch sensation. *Nature* **493,** 221–225 (2013).

31. Zhang, W., Yan, Z., Jan, L. Y. & Jan, Y. N. Sound response mediated by the TRP channels NOMPC, NANCHUNG, and INACTIVE in chordotonal organs of *Drosophila* larvae. *Proc. Natl Acad. Sci. USA* **110,** 13612–13617 (2013).

32. Robertson, J. L., Tsubouchi, A. & Tracey, W. D. Larval defense against attack from parasitoid wasps requires nociceptive neurons. *PLoS ONE* **8,** e78704 (2013).

33. Tracey, W. D. Jr, Wilson, R. I., Laurent, G. & Benzer, S. *painless*, a *Drosophila* gene essential for nociception. *Cell* **113,** 261–273 (2003).

34. Vogelstein, J. T. *et al.* Discovery of brainwide neural–behavioral maps via multiscale unsupervised structure learning. *Science* **344,** 386–392 (2014).

35. Xiang, Y. *et al.* Light-avoidance-mediating photoreceptors tile the Drosophila larval body wall. *Nature* **468,** 921–926 (2010).

36. Zhong, L. *et al.* Thermosensory and nonthermosensory isoforms of Drosophila melanogaster TRPA1 reveal heat-sensor domains of a thermoTRP Channel. *Cell Rep.* **1,** 43–55 (2012).

37. Schrader, S. & Merritt, D. J. Central projections of *Drosophila* sensory neurons in the transition from embryo to larva. *J. Comp. Neurol.* **425,** 34–44 (2000).

38. Zlatic, M., Li, F., Strigini, M., Grueber, W. & Bate, M. Positional cues in the *Drosophila* nerve cord: semaphorins pattern the dorso-ventral axis. *PLoS Biol.* **7,** e1000135 (2009).

39. Grueber, W. B. *et al.* Projections of *Drosophila* multidendritic neurons in the central nervous system: links with peripheral dendrite morphology. *Development* **134,** 55–64 (2007).

40. Klapoetke, N. C. *et al.* Independent optical excitation of distinct neural populations. *Nature Methods* **11,** 338–346 (2014).

41. Ernst, M. O. & Bulthoff, H. H. Merging the senses into a robust percept. *Trends Cogn. Sci.* **8,** 162–169 (2004).

42. Kenshalo, D. R., Iwata, K., Sholas, M. & Thomas, D. A. Response properties and organization of nociceptive neurons in area 1 of monkey primary somatosensory cortex. *J. Neurophysiol.* **84,** 719–729 (2000).

43. Mendell, L. M. Physiological properties of unmyelinated fiber projection to the spinal cord. *Exp. Neurol.* **16,** 316–332 (1966).

44. Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. A normalization model of multisensory integration. *Nature Neurosci.* **14,** 775–782 (2011).

45. Bargmann, C. I. & Marder, E. From the connectome to brain function. *Nature Methods* **10,** 483–490 (2013).

46. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans. Phil. Trans. R. Soc. Lond. B* **314,** 1–340 (1986).

47. Bock, D. D. *et al.* Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* **471,** 177–182 (2011).

48. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471,** 183–188 (2011).

49. Kim, J. S. *et al.* Space-time wiring specificity supports direction selectivity in the retina. *Nature* **509,** 331–336 (2014).

50. Takemura, S. Y. *et al.* A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500,** 175–181 (2013).

METHODS

Fly strains. In the main text and figures, short names are used to describe genotypes for clarity. The complete genotypes of animals used in this study are shown in Supplementary Table 2.

The following strains from the Rubin GAL4/LexA collection were used for the figures in the manuscript: *R61D08-GAL4, R69F06-GAL4, R72F11-GAL4, R20B01-GAL4, R71A10-GAL4, R57F07-GAL4, R47D07-GAL4, R16E11-LexA, R61D08-LexA, R72F11-LexA* and the control stock *w;;attp2* with no GAL4 or LexA[24,51]. For the thermogenetic activation screen we selected 586 GAL4 lines[26] that drove expression in relatively few neuron types (one to five) and used them to drive expression of TRPA with *UAS-dTRPA1*[52,53].

The following effector stocks were used: *pJFRC25-13xLexAop2-IVS-dTrpA1* in *vk05*[25,54]; *pJFRC48-13xLexAop2-IVS-mrytdTomato* in *attp18*, *pJFRC20-8xLexAop2-IVS-GAL80-WPRE* in *attp18*[25,54]; *pJFRC29-10xUAS-IVS-myr::GFP-p10* in *attp2*[54]; *pGP-JFRC7-20xUAS-IVS-GCaMP6s*[55] (Bloomington stock no. 42749); *pGP-JFRC7-20xUAS-IVS-GCaMP6f5* (Bloomington stock no. 52869); *20xUAS-IVS-GCaMP3* (Bloomington stock no. 32236;[56]); and *UAS-JFRC12-10xUAS-mryGFP* (Bloomington stock no. 32197;[51,54]) were gifts from B. D. Pfeiffer, G. Rubin, and the GECI project team (HHMI Janelia Research Campus). *ppk1.9-GAL4*[57] was a gift from D. Tracey (Duke University). *ppk-1kb-hs43-lexA-GAD* in attp40 was a stock that was previously described in ref. 34. *UAS-dTRPA1*[52] was a gift from P. Garrity. *UAS-EGFP-Kir2.1*[58] was outcrossed to *w1118* and gifted from D. Sitaraman (Yale University). *LexAop2-P2X2*[59] was a gift from C.-H. Yang (Duke University; unpublished stock).

For selective targeting the Goro neurons, we used *LexAop-FLP; UAS>stop>dTrpA1::myc*[60] a gift from B. Dickson, crossed with *w1118;16E11-LexA (attp40); R69F06-GAL4 (attp2)*, which drives expression in Goro neurons. The expression was determined by anti-Myc antibody (seen on immunohistochemistry).

Tsh-LexA was a gift from J.-M. Knapp and J. Simpson (unpublished stock). In brief, the *tsh-LexA* driver is an enhancer trap inserted into the 5′ UTR of the tsh locus. It was generated via a P-element swap that replaced the *p{GawB}* insertion of *tsh-GAL4*[61] with *P{UpP65L}* and the enhancer-trap LexA construct. Proper targeting and orientation of *P{UpP65L}* were confirmed by splinkerette PCR and sequencing by J.-M. Knapp (unpublished results).

LexAop2-mry-TDTomato-p10 (attp40), a gift from D. Mellert (HHMI Janelia Research Campus; unpublished stock), is an *mry::TDtomato* fragment with *AcNPV p10*[54]. To selectively target Basin-4 neurons we generated a Split-GAL4[25,62] stock *R72F11_AD* inserted in *attp40* (chromosome 2L) and *57F07_DBD* in *attp2* (3L). AD and DBD combinations were assembled in a *w1118* background. We identified *R57F07* as a candidate line for this intersection because its activation also evoked rolling in the thermogenetic activation screen (Extended Data Fig. 4). Stochastic labelling of single cells in *R57F07* and *R72F11*, using a FLP-based strategy (see below) revealed they both drive expression in Basin-4. *R57F07* also drove expression in other neurons in the tip of the nerve cord and brain (Extended Data Fig. 4).

The 'FLP-out' approach[63] for stochastic single-cell labelling will be described in detail elsewhere (Nern *et al.*, manuscript in preparation). In brief, heat-shock induced expression of FLP recombinase was used to excise FRT-flanked interruption cassettes from UAS reporter constructs carrying HA, V5, and Flag epitope tags, and stained with epitope-tag specific antibodies. This labelled a subset of the cells in the expression pattern with a stochastic combination of the three labels.

Behavioural apparatus. The apparatus comprises a video camera (DALSA Falcon 4M30 camera) for monitoring larvae, a ring light illuminator (Cree C503B-RCS-CW0Z0AA1 at 624 nm in the red), a computer (see ref. 28 for details; available upon request are the bill of materials, schematic diagrams and PCB CAM files for the assembly of the apparatus) and two hardware modules for controlling vibration and temperature (Oven Industries PA, Model 0805). Both hardware modules were controlled through multi worm tracker (MWT) software (http://sourceforge.net/projects/mwt), as described in ref. 28.

The module for presenting vibration consists of a speaker (120 W, 12 inches in diameter) and an amplifier (Pyle Pro PCA3) controlled by a signal generator (Tektronix AFG3021B). The speaker was used to generate vibrations between 100 and 1000 Hz, with an electronic signal (1 V, 100 Hz) generating a sound pressure level of 122 dB at the assay plate (measured by a Realistic 33-2050 Sound Level Meter). The frequency of vibration at the agar surface and vibration intensity (measured as acceleration in m s^{-2}) was measured with an ADXL345 3-axis accelerometer assembled onto a small printed circuit board placed on the agar as described in ref. 28.

For thermal activation, we used a Peltier module controlled by a temperature controller (Oven Industries PA, Model 0805). An aluminium plate covered with a thin layer of 4% charcoal agar was placed on the Peltier module. The temperature

of the agar was raised from 20 °C to 30 °C/32 °C to weakly/strongly activate the dTRPA1 channel. By using a temperature gun, we confirmed that the agar reached the Peltier tile setting temperature. The time to reach the target temperature ranged from 3 to 8 s. The behaviour rigs were inside 32.00" wide × 28.00" deep × 60.00" high temperature-controlled enclosures with temperature settable from ambient to 40 °C in 0.1 °C steps (Life Science Engineering, Inc.). The humidity in the room was monitored and held at 58%, with humidifiers (Humidifirst Mist Pac-5 Ultrasonic Humidifier).

Behavioural experiments. Embryos were collected for 6 h at 18 °C and 65% humidity. Larvae containing the *UAS-dTrpA1* or *LexA-dTrpA1* transgene were raised at 18 °C for 7 days with normal cornmeal food. Foraging third instar larvae were used for all experiments. Larvae containing the *UAS-CsChrimson* transgene were grown in the dark at 25 °C for 4 days on fly food containing *trans*-retinal (Sigma, R2500) at a concentration of 500 M.

Before the experiments, the larvae were separated from food by using 15% sucrose and washed with water. The larvae were then dried and placed in the center of the arena. The substrate for the behavioural experiments was 3% Bacto agar gel in a 25 × 25 cm square plastic plate for experiments involving thermal activation and vibration stimuli, or a 10 × 10 cm plate for those involving thermal activation alone. We tested approximately 15–50 larvae at once in the behavioural assays. The temperature of the entire rig was kept at 25 °C (for optogenetic activation experiments) or 30 °C or 32 °C for thermogenetic activation experiments. The agar plates were also kept at the desired temperature prior to experiment. For optogenetic activation experiments we exposed dishes of larvae to red light (630 nm, 2.4 mW per cm^2). Optogenetic activation offers greater temporal precision than thermogenetic activation so we used to study the ability of Basin, Goro and A00c neurons to trigger the entire behavioural sequence roll-fast crawl, or just a single element of the sequence (roll or crawl)

The MWT software[64] (http://sourceforge.net/projects/mwt) was used to record all behavioural responses and to control the presentation of vibration stimuli.

For the thermogenetic activation screen we used 586 GAL4 lines preselected for expression in relatively few neuron types (one to five) to drive expression of dTRPA1. In each experiment, we exposed dishes of larvae to 32 °C to thermogenetically activate dTRPA1-expressing neurons. We detected 32 'hits' that contained neurons whose thermogenetic activation evoked rolling in a significant fraction of animals. We analysed further hits that evoked rolling in the highest fraction of larvae and that drove expression in the least neuron types.

Behaviour quantification. Larvae were tracked in real-time using the MWT software. We rejected objects that were tracked for less than 5 seconds or moved less than one body length of the larva. For each larva MWT returns a contour, spine and centre of mass as a function of time. Raw videos are never stored. From the MWT tracking data we computed the key parameters of larval motion, using specific choreography (part of the MWT software package) variables[28]. From the tracking data, we detected and quantified crawling and rolling events and the speed of peristaltic crawling strides as described in ref. 28, using the LARA software package (http://sourceforge.net/projects/salam-hhmi). We also trained a machine-learning classifier, using JAABA (Janelia Automatic Animal Behavior Annotator[65]) to detect rolling events from the tracking data. The false detection rate of the JAABA rolling classifier was slightly smaller (2.8%, *n* = 89) than that of the LARA detection software (3.6%), but the results obtained with both methods were very similar. The MATLAB code for the rolling classifier is available upon request. The data presented is a quantification of rolling events detected with JAABA. To calculate the percentage of animals that rolled in response to thermogenetic activation, vibration, or both, we selected for analysis all the animals that were tracked throughout the entire 15-second sampling interval following stimulation. We then calculated the percentage of animals from this sample for which at least one rolling event was detected during the 15-second interval. The instantaneous percentage of rolling and crawling larvae (Extended Data Fig. 7a) as a function of time was calculated by dividing the number of larvae performing rolling at each time point with the total number of larvae tracked at each time point.

Selective targeting of Goro neurons. We expressed FLP selectively in the Goro neurons using the weak 16E11 LexA lines (in *16ElexA/LexAop FLP; UAS>stop>UAS-dTRPA1::myc/R69F06*). Selectively in Goro neurons, FLP stochastically excises the STOP cassette that prevents the expression of *UAS-dTRPA1*. Now the strong *R69F06* promoter can drive high levels of dTRPA1 expression, selectively in Goro neurons. Since this method of Goro activation is stochastic we monitored rolling in response to 32 °C exposure in individual larvae. We dissected all animals, noted whether they rolled or did not, and stained their nervous systems with anti-Myc, to identify those that had Goro expression.

Statistical analysis. Statistical analysis was performed using MATLAB (MathWorks) software. The calcium responses in the paired trials shown in Fig. 4f were compared both using the paired *t*-test and Wilcoxon signed-rank

test. Required sample size for power 0.8 was determined using GPower software for paired *t*-test and for a large pre-specified effect size based on a few pilot trials. Indeed the effect size in the data was large ($P = 0.016$ and $d = 1.05$ for one-sided paired *t*-test and $P = 0.02$ and $r = 0.81$ using Wilcoxon signed-rank test). The speed of peristaltic crawling strides was compared using Wilcoxcon rank sum test (and Bonferroni correction for multiple comparisons). In all figures, * and ** represent $P < 0.05$ and $P < 0.01$, respectively. Required sample size for power 0.8 was determined using GPower software for a medium pre-specified effect size. The fractions of animals that rolled under the different experimental conditions were compared using the χ^2 test. *P* values were corrected for multiple comparisons using Bonferroni correction. Power analysis was performed as described below.

We performed an analysis of the power of our statistical tests of behavior differences, given the sample sizes and effect sizes observed. In addition, we computed the minimum effect sizes our analyses would have been sensitive to with a power 0.95, given the sample sizes. In our power analyses, we assumed that the distribution of the fraction of larvae that rolled for the control populations, q_0, was known and equal to the fraction measured in a large number of trials.

For each experiment, we tested the null hypothesis that the fraction of experimental animals that rolled was less, more, or equal to the control fraction, depending on the type of experiment. For example, when activating the Goro neurons, we tested whether we could reject the null hypothesis that the fraction of larvae that rolled was less than or equal to the control fraction. In this case, assuming that each animal is an independent sample from the same distribution, the number of larvae rolling follows a binomial distribution. Thus, if the null hypothesis is that the experimental larvae roll at most as much as control, the probability that the null hypothesis holds can be computed from the binomial cumulative distribution function:

$$\Pr_{\leq q_0}(n,\hat{q}n) = 1 - \text{binocdf}(\hat{q}n, q_0)$$

Where n is the number of experimental larvae tested, $\hat{q}$ is the measured fraction that roll, and binocdf (k,r) is the binomial cumulative distribution function. We assumed we would reject the null hypothesis if this probability was less than 0.01. For a given q_0 and n, we can thus compute $\hat{q}^*_{q_0,n}$, the minimum $\hat{q}$ such that we would reject the null hypothesis using this equation.

To compute the power of this statistical test given the sample size n and true fraction of experimental larvae expected to roll q, we assumed that our collected experimental data were independent draws from the Bernoulli distribution with parameter q, and computed the probability that our observed fraction of

$$\Pr(\text{reject null hypothesis}|q_0 \leq q, n) = 1 - \text{binocdf}\left(\hat{q}^*_{q_0,n}n, q_0\right)$$

Analogous equations were used for testing whether the experimental rolled at least as much as control or the same amount as control.

Using these equations, we computed the power of each statistical test given the observed sample size n, control fraction that rolled q_0, and the experimental fraction that rolled q, for each set of experiments. We also computed the minimum effect size q / q_0 that would be sensitive with power 0.95, given the observed sample size n and control fraction that rolled q_0. These results are listed in Supplementary Tables 1 and 3.

Immunohistochemistry. The following antibodies were used: rabbit anti-GFP (1:1000, A11122, Invitrogen), rat anti-N-Cad (1:50, Developmental Studies Hybridoma Bank), mouse anti-nc82 (1:20; Developmental Studies Hybridoma Bank), mouse-anti-Myc (9E10, 1:20, Developmental Studies Hybridoma Bank), rabbit anti-HA-Tag (1:300, Cell Signaling, #3724), rat anti-FLAG (1:200, NOVUS, NPP1-06712), mouse anti-V5 Dylight549 (1:400, AbD Serotec, MCA1360D549). For secondary antibody Alexa Fluor 488 goat anti-rabbit IgG (1:200, Invitrogen), Alexa Fluor 568 goat anti-mouse IgG (1:200, Invitrogen) and Alexa Fluor 488 goat anti-mouse IgG (1:200, Invitrogen) were used.

The central nervous systems of wandering third instar larvae were dissected in cold Schneiders Insect medium (Sigma, S0146) and fixed in 4% phosphate-buffered saline (PBS) with 4% paraformaldehyde (Electron Microscopy Science) at room temperature for 20 min. After washes in 0.4% Triton X-100 in PBS, samples were preblocked with 3% normal goat serum (NGS) in for 30 min at room temperature. The samples were then incubated in primary antibody for 2 to 3 days at 4 °C. After extensive washes in 0.4% Triton X-100 in PBS, the samples were incubated at 4 °C for 2 days in a secondary antibody solution. After further extensive washes, the samples were rinsed in PBS, dehydrated through an EtOH series, cleared in Xylene and mounted in DPX mountant (Sigma-Aldrich). Images were taken with a 710 or 700 laser-scanning confocal microscope (Zeiss) under 20× 0.8 NA or 63×, 1.4 NA objective and processed in ImageJ software[66] (NIH).

Functional connectivity assay (GCaMP recording). For activation of presynaptic neurons (MD IV, chordotonal, Basins or A05q) with the ATP-gated $P2X_2$ channel and imaging in postsynaptic neurons (Basins, Goro, A00c or A05q), the central nervous system of wandering third instar larvae were dissected out in cold physiological saline containing 103 mM NaCl, 5 mM KCl, 5 mM HEPES, 26 mM $NaHCO_3$, 1 mM NaH_2PO_4, 5 mM Trehalose, 6 mM sucrose, 2 mM $CaCl_2$ 2 mM H_2O, 8 mM $MgCl_2$ 6 mM H_2O, and kept stable by sticking them on poly-L-lysine (SIGMA, P1524) coated cover glass placed in small Sylgrd (Dow Corning) plates. A borosilicate pipette (1 to 2 µm tip diameter) filled with 5 mM adenosine 5′-triphosphate (ATP, Amersham 27-1006) and 50 g ml^{-1} Alexa Fluor 594 hydrazide (Invitrogen, A10438) dye in physiological saline was positioned 3 to 5 µm from the axon terminals of the target presynaptic neurons expressing the ATP-gated $P2X_2$ channel and dsRed and the ATP solution was pressure-ejected using a Picospritzer (Parker Hannifin Co.).

When ejecting ATP onto the axon terminals of MD IV or chordotonal neurons, we imaged calcium transients in the dendritic arbor of the Basin neurons, not the cell bodies (which in the insect are removed from the neuropil). When stimulating Basins in abdominal segments, by ejecting ATP onto their axon terminals in ventral nerve cord, we imaged calcium transients in the dendrites of Goro (in dorsal thoracic nerve cord), or in axon terminals of A05q (in dorsal nerve cord) and A00c (in the brain). When stimulating A05q by ejecting ATP onto their dendrites in ventral nerve cord in abdomen, we imaged calcium transients in Goro dendrites in dorsal thoracic nerve cord.

Puffs were 25 ms in duration, and the pressure was adjusted so that red fluorescence (from the Alexa 594) ejection could be seen clearly while imaging. The same stimulus was presented every 30 sec for a total of three trials in each experiment.

When vibration stimuli were also used in experiments, rather than only $P2X_2$ activation, the larvae were dissected using methods similar to those described previously for neuromuscular junction[67] on the Sylgard plate. In brief, larvae were cut up along the dorsal body wall midline and then opened by placing dissection pins. The cell bodies of abdominal peripheral sensory neurons and their nerves were kept intact. The anterior tips were removed, and the CNS was attached to a poly-L-lysine coated cover glass. For vibration stimuli in the two-photon rig, we embedded piezo extension actuators (Piezo System Inc., Model Q220-A4-303XE) in the Sylgard plate and used TTL pulses from the Prairie View software to control the timing of stimulation. The acceleration was measured with an ADXL345 3-axis accelerometer[28]. The vibration stimuli were 1,000 Hz and 6.67 m s^{-2}, 15 sec long. The same stimulus was presented every 30 sec for a total of three trials in each experiment. For comparing the responses of Basin-4 to combined activation (nociceptor activation with $P2X_2$ and ATP injection plus vibration) to nociceptor activation or vibration alone (Figure 4d and e), each trial consisted of ATP injection, vibration, and a combination (ATP injection and vibration), with 30 sec between each stimulation. Three such triplet trials were presented in each experiment.

We performed imaging on a Prairie Technologies two-photon microscope using an Olympus 60× 0.8 NA objective. A mode-locked Ti:Sapphire Chameleon Ultra II laser (Coherent) tuned to 925 nm was used for excitation. Fluorescence was collected with photomultiplier tubes (Hamamatsu) after band-pass filtering. Images were acquired in frame-scan mode (maximum speed, 1,215 f.p.s.) for a single plane of the CNS.

GCaMP image analysis. Image data were processed by ImageJ software[66] (NIH) and analysed using custom code written in MATLAB (MathWorks). Specifically, regions of interest (ROIs) were determined by averaging the ten frames before stimulation and segmenting this data by the function MEAN[68] in ImageJ. The mean intensity of the ROI was measured in ImageJ. In all cases, changes in fluorescence were calculated relative to baseline fluorescence levels (F_0) as determined by averaging over a period of at least 3 sec just before ATP puff or vibration stimulation. $\Delta F/F_0$ values were calculated as $\Delta F/F_0 = (F_t - F_0)/F_0$, where F_t is the fluorescent mean value of a ROI in a given flame. Peak $\Delta F/F_0$ is the maximum value in the 5 sec window following stimulus onset.

Electron microscopy preparation and imaging. For generating the electron microscopy dataset comprising 1.5 abdominal segments (segment a2 and a3), the central nervous systems from 12–24-h-old wild-type w^{1118} first instar larvae were manually dissected out in PBS, and immediately transferred to 2% glutaraldehyde in 0.1 M sodium cacodylate, pH 7.4 buffer. Samples were post-fixed in 1% OsO_4 in the same buffer and stained en bloc with 1% aqueous uranyl acetate before subsequent dehydration in ethanol and propylene oxide, and embedding in Epon. Serial 45 nm sections were cut with a Leica UC6 ultramicrotome using a Diatome diamond knife, and picked up with Synaptek slot grids with Pioloform support films. Sections were stained with uranyl acetate followed by Sato's lead[69].

For generating the electron microscopy dataset comprising the entire central nervous system, central nervous systems from 6-h-old *[iso] Canton S G1 x w^{1118}*

[iso] 5905 female larvae were manually dissected out in PBS. The isolated central nervous systems were immediately transferred to 125 μl of 2% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4 in a 0.5 dram glass vial (Electron Microscopy Sciences, cat. no. 72630-05) to which an equal volume of 2% OsO_4 in 0.1 M sodium cacodylate buffer, pH 7.4 was added and gently mixed. Each CNS was then fixed in an ice bath in a Pelco BioWave PRO microwave oven (Ted Pella, Inc.) at 350-W, 375-W and 400-W pulses for 30 sec each, separated by 60-sec intervals with the microwaves off. Samples were rinsed 3 × 30 sec at 350 W with 0.1 M sodium cacodylate buffer, separated by 60-sec intervals with the microwaves off, and post-fixed with 1% OsO_4 in 0.1 M sodium cacodylate buffer at 350-W, 375-W and 400-W pulses for 30 sec each, separated by 60-sec intervals with the microwaves off. After rinsing with distilled water 3 × 30 sec at 350 W with 60-sec pauses between pulses, the samples were stained en bloc with 1% uranyl acetate in water by microwave at 350 W for 3 × 30 sec with 60-sec pauses between microwave pulses. The samples were then dehydrated in an ethanol series followed by propylene oxide, infiltrated and embedded with EPON resin. Serial 50-nm sections were cut with a Leica UC6 ultramicrotome using a Diatome diamond knife, and picked up with Synaptek slot grids with Pioloform support films with 2 nm of carbon (C). Sections were stained with uranyl acetate followed by Sato's lead[69].

Both data sets were imaged using Leginon[70], the 1.5 abdominal segment on an FEI T20 TEM (Hillsboro) at 4.4 nm × 4.4 nm pixel resolution, the complete L1 brain on an FEI Spirit TEM at 3.8 nm × 3.8 nm pixel resolution. The resulting images were montaged and registered using the nonlinear elastic method described in ref. 23.

Neuronal reconstruction. Reconstructions were made in a modified version of CATMAID[71] (http://www.catmaid.org). The process of reconstructing neurons in CATMAID involves starting with a specified neurite in a section of the electron microscopy volume, and manually building a 3D skeleton representation of the neuronal morphology and the location of synaptic active zones and synaptic partners.

The synaptic connections mapped and reported in this study all represent fast, chemical synapses. Synapses must match several criteria over multiple adjacent sections: a thick black active zone, vesicles, and presynaptic specializations, such as a T-bar, on the presynaptic side, and evidence of postsynaptic membrane specializations[72].

To identify sensory axons, reconstruction was initiated at the entry point of nerves into the neuropil. Sensory neurons project ventrally in the nerve cord, and motor neurons project dorsally[37,38,73], so the sensory axons were readily identified in the ventral portion of the nerve root. All sensory axons were traced and MD IV and chordotonal axons were recognized based on their distinguishing features, previously described using confocal microscopy imaging[37,38,74]. Chordotonal projections were readily identified based on the characteristic intermediate location of their terminals in the mediolateral axis and the dorsoventral axis of the nervous system. We found that all eight chordotonal axons were already tightly bundled together in the nerve root. The MD IV projections were readily recognized based on the characteristic medial-most and ventral-most location of their terminals. Neuronal partners of the eight chordotonal axons and the three MD IV axons in both hemisegments were reconstructed by initiating reconstructions at the site of every postsynaptic target in all chordotonal terminals until all targets were exhausted. A review of all skeletons was done by CMSM to identify incorrect or missed branches and synapses.

In the comprehensive L1 electron microscopy volume, Basin cells in A1 and A4 segments were found using landmarks such as cell body position, the entry point of primary neurite into the neuropil, and branch points on the primary neurite. In brief, the Basin lineage was first recognized on the basis of the characteristic cell body location (lateral intermediate), the trajectory of their primary neurite which is parallel to the entering nerve, and the neuropil entry point of the lineage, as well as the pattern of the low-order, large branches. The lineage contains more cells than just the four Basins, but the other cells have very different projection patterns. The Basins in the A1 segment were identified by tracing the primary and secondary neurites of multiple members of the lineage and abandoning those that did not have the distinguishing features of the Basins. The four Basins in each hemisegment were reconstructed in full and individual Basins recognized based on their unique features (see Extended Data Fig. 2). Basin targets were reconstructed by tracing exhaustively downstream of all presynaptic sites in Basin axons. Goro neurons were found in a similar way as Basins. Starting with high-resolution 3D image stacks of single cells (visualized using the FLP-based strategy described above), we looked at the cell body location, primary neurite projection, and major branch points. We then traced neurites consistent with the primary neurite projection and cell body location until finding only two individual cells, one in each hemisegment of segment T2, with branch points consistent with Goro neurons. Completed reconstructions were then checked for consistency with light

data in axon and dendritic projections. Partial reconstructions of sibling neurons, as well as homologous neurons in adjacent thoracic segments, resulted in morphologies very different from that of Goro neurons as observed by light microscopy, indicating that only two neurons in the whole of the thorax match the expected morphological pattern. Descending neurons were found by reconstruction of neurons upstream of synaptic inputs in Goro dendrites. A subset of fourth-order neurons was found by reconstructing downstream of all synaptic output sites of only one of the ascending third-order A00c-a4 neurons. Left and right homologues of reconstructed neurons were identified by considering a suite of morphological and synaptic features, focusing on neuropil entry location, primary branch locations, and the location of pre- and postsynaptic sites.

Neuronal morphologies were analysed in MATLAB using custom scripts. Circuit diagrams were made in Gephi[75] and edited in Adobe Illustrator CS6 (Adobe Systems, Inc.).

Validation of methodology of electron microscopy reconstruction. In *Drosophila*, as in other insects, the gross morphology of many neurons is stereotyped and individual neurons are uniquely identifiable based on morphology[37,50,76,77]. Furthermore, the nervous system in insects is largely bilaterally symmetric and homologous neurons are reproducibly found on the left and the right side of the animal. In this study we therefore validate the wiring diagram by: (1) independently reconstructing partners of homologous neurons on the left and right side of the nervous system; (2) independently reconstructing partners of homologous neurons in EM volumes from two different individuals; (3) performing functional connectivity experiments to validate some key connections predicted by the wiring diagram.

Validating the wiring diagram by independent reconstruction on the left and right. By randomly re-reviewing annotated synapses and terminal arbors in our dataset we estimated the false positive rate of synaptic contact detection to be 0.0167 (1 error per 60 synaptic contacts). Assuming the false positives are uncorrelated, for an *n*-synapse connection the probability that all *n* are wrong (and thus that the entire connection is a false positive) occurs at a rate of 0.0167^n. Thus, the probability that a connection is false positive reduces dramatically with the number of synaptic connections contributing to that connection.

Even for $n = 2$ synapse connections, the probability that the connection is not true is 0.00028 (once in every 3,586 two-synapse connections) and we call connections with two or more connections 'reliable' connections. Similarly, the probability that even a one-synapse connection annotated between homologous neurons both on the left and on the right side of the nervous system is not true is 0.00028. The probability that a false-positive connection with two synapses is observed twice between the same cell types is even lower (0.000282). Because false-positive synapses between a single pair of homologous neurons could potentially be correlated (due to mis-annotation of a single branch with multiple contacts), we perform extra rounds of focused re-review of the arbors associated with connections in our wiring diagram with less than four synapses to confirm that they are neither mis-annotated synapses or incorrect branches. Furthermore, while false positives between a single pair of homologous neurons could potentially be correlated, the false positives between different pairs of homologous neurons (on the left side and the right side) cannot be correlated (they are not on the same branch). Thus the probability that reliable (with two or more synapses) and reproducible (found between both the left and right homologues within an individual) connections are not true is extremely low and we are highly confident in such connections. All the connections discussed in this study and all the principles derived from the wiring diagram are based on such reliable and reproducible connections.

False negatives are far more common in our reconstructions than false positives. We were able to trace 81.5% (1,676/2,054) of postsynaptic sites downstream of Basins back to identifiable neuronal arbors, leaving 18.5% (378/2,054) of synaptic outputs associated with small, unidentifiable arbor fragments due to ambiguities in the data. The majority have only a single synaptic input from Basins (163/231 fragments), and only 8/231 fragments have five or more Basin inputs. If the synapses that could not be associated with identified neurons were distributed randomly, the likelihood of completely omitting a connection with *n* synapses is approximately 0.185^n, which is 0.006 for $n = 3$. Thus the likelihood that we have completely omitted a three-or-more-synapse connection is quite low.

Reproducibility of the wiring diagram across different individuals. For a connection in the wiring diagram to be useful for future physiological and behavioural experiments in different individuals, it should not just be correctly identified in one animal, but it should be generalizable across animals. We therefore assessed the extent to which the connections in this wiring diagram are reproducible across individuals.

For this study we obtained electron microscopy volumes from two different larvae: one comprising 1.5 nerve cord segments (a2/a3), and one comprising the

entire nervous system. We reconstructed the chordotonal and MD IV sensory neurons, the Basin neurons and all the other partners downstream of both the nociceptive and the sensory neurons in both electron microscopy volumes. In the smaller abd1.5 volume some of the downstream neurons had a large fraction of their arbor outside the volume and were not identifiable, whereas others were mostly confined within the electron microscopy volume and were uniquely identifiable based on key morphological features. We could therefore assess the reproducibility of connections between homologous neurons present and uniquely identifiable in both volumes across two individuals (Extended Data Fig. 3). We identified connections between neuron types that have at least two synapses and for which a connection with at least two synapses also exists between homologous neuron types on the other side of the same individual. 100% ($n = 36$) of connections with at least two synapses per connection that were reproducible across the same individual were also reproducible across individuals; that is, those connections also existed on both sides in the other individual.

If we consider each individual sensory neuron in a hemisegment as unique, we find a total of 76 connections (38 pairs of connections) between individual neurons with at least two synapses between homologues on both the left and right of the same individual. In 100% of cases in which a connection was present between homologues on both sides in one animal it was also present between homologues in the other animal, at least on one side. 96% of connections with at least two synapses that existed between homologous neurons on both sides in one animal also existed between homologues on both sides in the other animal. 4% of connections with at least two synapses present between homologues on both sides in one animal were not present between homologues in the other animal on the same side, but a connection was present between homologues on the other side (in the other hemisegment) in the other individual (see Extended Data Fig. 3 for more details).

Together with the above error rates, the observed reproducibility of connections across individuals gives us strong confidence that the observation of a reliable connection in one animal both measures a real connection in that individual and predicts similar connectivity in other animals.

Functional validation of connections mapped using electron microscopy reconstruction. Another way to validate a wiring diagram is to functionally test the connections obtained with electron microscopy reconstruction. If the anatomical and functional connectivity studies are done in different individuals this provides further validation of the reproducibility and generalizability of the wiring diagram. In this study we were able to generate reagents to functionally test a number of anatomical connections predicted by the wiring diagram. By activating the presynaptic partner and monitoring calcium transients in the direct postsynaptic partner we tested the following reproducible direct anatomical connections with, on average, five or more synapses per connection:

• MD IV onto Basin-4 (average >10 synapses per connection) (Fig. 3b)

• Chordotonal onto Basins (average >10 synapses per connection) (Extended Data Fig. 4d)

• Basins onto A00c (average >10 synapses per connection) (Extended Data Fig. 6j)

• Basins onto A05q (average >5 synapses per connection) (Extended Data Fig. 6e)

• A05q onto Goro (average 5 synapses per connection) (Extended Data Fig. 6g)

For the anatomical connection from MD IV onto Basin-1 with only two synapses per connection on average, the functional connectivity experiments did not reveal a functional excitatory connection (Fig. 3a).

Model of multilevel convergence. The model consists of two sensory inputs processed in two layers (Extended Data Fig. 8a). The sensory inputs, chordotonal and MD IV, have activity E_{Mech} and E_{Noci} respectively. Sensory input intensity can range from 0–100 (arbitrary units), with the maximum value intended to approximate the highest intensity stimulus provided in the behavioural experiments of Fig. 1. The first processing layer has two nodes, each representing one or more neurons: (1) a 'pure' node with output $g_P(E_{Mech})$ that receives only mechanosensory input; and (2) a 'mixed' node with output $g_M(E_{Mech}, E_{Noci})$ that receives input from both sensory modalities. The second processing layer is a single node with output $g_B(g_P, g_M)$ that receives input from both nodes of the first processing layer. The output of the second layer constitutes the output of the circuit.

We considered the steady-state output of the network, assuming that the response of each node is a sigmoidal function of its total input, with a lower threshold below which output is near-zero and an upper saturation point. We modelled this as a logistic function[78]:

$$f(E; \theta, \beta) = C_{\theta, \beta} \frac{1}{1 + e^{-\frac{(E-\theta)}{\beta}}}$$

with total input E, threshold θ, transition width β, and a constant $C_{\theta, \beta}$ defined so that $f(100) = 100$. We set $\beta = 10$ for all nodes, which makes the transition between

low and high output cover about half of the input range. We omit β in subsequent equations for brevity.

The output of the pure node in the first processing layer is given by: $g_P(E_{Mech}) = f(E_{Mech}; \theta_P)$.

For the mixed node, we take the total input to be a weighted sum of the sensory inputs: $g_M(E_{Mech}, E_{Noci}) = f(w_M E_{Mech} + (1 - w_M)E_{Noci}; \theta_M)$ where $0 \le w_M \le 1$ measures the relative weighting of mechanosensory input. To require that the maximum total input each node can receive is the same, the weights sum to one.

For the second layer, we take total input to be a weighted sum of the outputs from the nodes in the first layer:

$$g_B(E_{Mech}, E_{Noci}) = f(w_B g_P(E_{Mech}) + (1 - w_B)g_M(E_{Mech}, E_{Noci}); \theta_B)$$

where $0 \le w_B \le 1$ measures the weighting of the pure chordotonal node in the first layer.

A common feature of multisensory circuits is their enhanced sensitivity to combinations of weak stimuli presented together. To consider this property in the context of our circuit, we looked for circuit responses that are maximally sensitive to multisensory stimuli, but that are also consistent with basic experimental observations. We treated this as a constrained optimization problem; we found mixing parameters w_M and w_B that optimize a measure of multisensory sensitivity, subject to the constraint that the responses to unimodal stimuli are similar to experimental observations (Extended Data Fig. 8b–d). We looked at the resulting solutions for a range of threshold parameters.

For the sensitivity measure, we assumed that the optimal solutions are responsive to combinations of both inputs (Extended Data Fig. 8b), as measured by the integrated response along a Gaussian-weighted band centred along the $E_{Mech} = E_{Noci}$ diagonal line:

$$S(w_M, w_B) = \int_0^{100} \int_0^{100} g_B(E_{Mech}, E_{Noci})e^{-(E_{Mech} - E_{Noci})/4d^2} dE_{Mech} dE_{Noci}$$

where d measures the width of the band, which we set to 20 (Extended Data Fig. 8c). Although this condition alone would encourage strong responses to extremely low values of input stimulus, an undesirable situation, the combination of the constraint and the monotonicity of the response functions ensure that this cannot occur.

For the constraint, we demanded that intense unimodal input match observations (see Fig. 1c): (1) that intense, pure chordotonal input induced no rolling; and (2) intense, pure MD IV input induced a modest, but non-zero, fraction of animals to roll (Extended Data Fig. 8d). We quantified this as:

$$h(w_M, w_B) = \sqrt{(g_B(0, 100; w_M, w_B) - a_{Noci})^2 + (g_B(100, 0; w_M, w_B) - a_{Mech})^2}$$

where we set $a_{Noci} = 30$ and $a_{Mech} = 0$ (Extended Data Fig. 8e).

We then calculated optimal connection weights given neuron thresholds. We assumed that the pure mechanosensory interneuron does not have an extremely high threshold, set θ_P to a value between 30 and 50 and explored the constrained optimization problem for θ_M and θ_B values between 10 and 80. For each set of parameters, the multilevel convergence solution was defined as the circuit output corresponding to the parameters (w_M, w_B) that maximize $S(w_M, w_B)$ subject to $h(w_M, w_B) < h_0$. We similarly defined the early convergence solution as the circuit with the w_M that maximizes $S(w_M, 0)$ subject to $h(w_M, 0) < h_0$. We set $h_0 = 3$ to ensure strong compliance with our unimodal constraints and numerically computed and compared solutions across a grid of w_B and w_B values starting at 0 with 0.01 spacing (Extended Data Fig. 8f). To check the robustness of early and multilevel convergence solutions to threshold parameters, we explored multiple values of θ_B and θ_B (Extended Data Fig. 8g).

51. Jenett, A. et al. A GAL4-driver line resource for Drosophila neurobiology. Cell Rep. **2**, 991–1001 (2012).
52. Hamada, F. N. et al. An internal thermal sensor controlling temperature preference in Drosophila. Nature **454**, 217–220 (2008).
53. Kang, K. et al. Modulation of TRPA1 thermal sensitivity enables sensory discrimination in Drosophila. Nature **481**, 76–80 (2012).
54. Pfeiffer, B. D., Truman, J. W. & Rubin, G. M. Using translational enhancers to increase transgene expression in Drosophila. Proc. Natl Acad. Sci. USA **109**, 6626–6631 (2012).
55. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature **499**, 295–300 (2013).
56. Tian, L. et al. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. Nature Methods **6**, 875–881 (2009).

57. Ainsley, J. A. *et al.* Enhanced locomotion caused by loss of the *Drosophila* DEG/ENaC protein Pick-pocket1. *Curr. Biol.* **13,** 1557–1563 (2003).
58. Baines, R. A., Uhler, J. P., Thompson, A., Sweeney, S. T. & Bate, M. Altered electrical properties in *Drosophila* neurons developing without synaptic transmission. *J. Neurosci.* **21,** 1523–1531 (2001).
59. Lima, S. Q. & Miesenböck, G. Remote control of behavior through genetically targeted photostimulation of neurons. *Cell* **121,** 141–152 (2005).
60. von Philipsborn, A. C. *et al.* Neuronal control of *Drosophila* courtship song. *Neuron* **69,** 509–522 (2011).
61. Calleja, M., Moreno, E., Pelaz, S. & Morata, G. Visualization of gene expression in living adult *Drosophila. Science* **274,** 252–255 (1996).
62. Luan, H., Peabody, N. C., Vinson, C. R. & White, B. H. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron* **52,** 425–436 (2006).
63. Struhl, G. & Basler, K. Organizing activity of wingless protein in *Drosophila. Cell* **72,** 527–540 (1993).
64. Swierczek, N. A., Giles, A. C., Rankin, C. H. & Kerr, R. A. High-throughput behavioral analysis in *C. elegans. Nature Methods* **8,** 592–598 (2011).
65. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nature Methods* **10,** 64–67 (2013).
66. Schneider, C. A. *et al.* NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9,** 671–675 (2012).
67. Verstreken, P., Ohyama, T. & Bellen, H. J. in *Exocytosis and Endocytosis* 349–369 (Springer, 2008).
68. Glasbey, C. A. An analysis of histogram-based thresholding algorithms. CVGIP. *Graph. Models Image Proc.* **55,** 532–537 (1993).
69. Sato, T. A modified method for lead staining of thin sections. *J. Electron Microsc. (Tokyo)* **17,** 158–159 (1968).
70. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151,** 41–60 (2005).
71. Saalfeld, S., Cardona, A., Hartenstein, V. & Tomancak, P. CATMAID: collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* **25,** 1984–1986 (2009).
72. Prokop, A. & Meinertzhagen, I. A. Development and structure of synaptic contacts in *Drosophila. Semin. Cell Dev. Biol.* **17,** 20–30 (2006).
73. Landgraf, M., Snchez Soriano, N., Technau, G. M., Urban, J. & Prokop, A. Charting the *Drosophila* neuropile: a strategy for the standardised characterization of genetically amenable neurites. *Dev. Biol.* **260,** 207–225 (2003).
74. Grueber, W. B., Jan, L. & Jan, Y. Tiling of the *Drosophila* epidermis by multidendritic sensory neurons. *Development* **129,** 2867–2878 (2002).
75. Bastian, M. *et al.* Gephi: an open source software for exploring and manipulating networks (International AAAI Conference on Weblogs and Social Media, 2009).
76. Bate, M., Goodman, C. & Spitzer, N. Embryonic development of identified neurons: Segment-specific differences in the H cell homologues. *J. Neurosci.* **1,** 103–106 (1981).
77. Jefferis, G. S. *et al.* Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* **128,** 1187–1203 (2007).
78. Dayan, P. & Abbott, L. F. *Theoretical Neuroscience* (MIT Press, 2001).
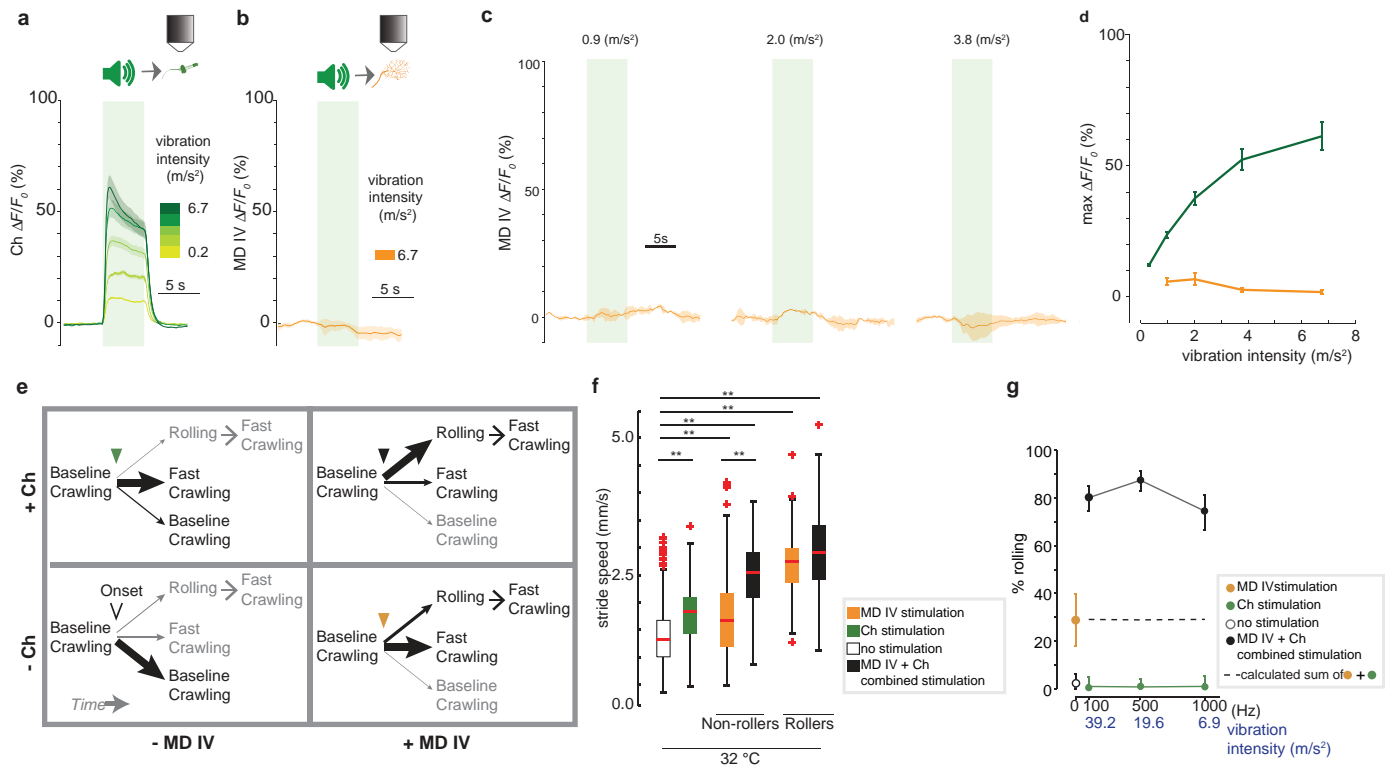
Extended Data Figure 1 | This figure relates to Fig. 1a–d. Chordotonal (Ch) neurons respond to vibration in a 'dose-dependent' manner, but MD IV neurons do not, consistent with previous reports. **a**, 1,000 Hz vibration evokes calcium transients (mean ± s.e.m. in this and all subsequent figures) in chordotonal (*ch-GAL4>UAS-GCaMP6s*), in a dose-dependent manner. **b**, **c**, Calcium transients of MD IV neurons in response to different vibration (1,000 Hz) intensities reveal no response to vibration (*MD IV-GAL4>UAS-GCaMP6s*). **d**, Quantification of calcium transients in mechanosensory (black) and nociceptive (orange) neurons from (**a–c**). $n = 3$ trials in each condition. **e**, A schematic of larval reactions to unimodal and multimodal stimuli. Line thickness qualitatively reflects the probability of the behaviour transition occurring in the absence of stimulation (bottom left) and after the onset of MD IV activation (orange triangle, bottom right), vibration (green triangle, top left), or their combination (black triangle, top right). Behaviours in grey occur with extremely low frequency. In response to nociceptive or combined stimulation animals can select one of two types of escapes responses: (1) roll-fast crawl; or (2) fast crawl alone. Even in response to combined stimulation the selection is probabilistic; the majority of animals select roll-fast crawl, but some select fast crawl without rolling. **f**, Crawling stride speed in the absence of stimulation (white) and following 1,000 Hz, 6.7 m s^{-2} vibration (green) (in *MD IV-GAL4>+* larvae); thermogenetic nociceptive activation (32 °C, in *MD IV-*

GAL4>UAS-dTRPA1 larvae) (orange); or their combination (grey). In the nociceptive and combined condition, we separately analysed the stride speed of animals that did not roll (non-rollers) and those that rolled (rollers) following stimulation. As previously described, vibration alone evoked an increase in stride speed. Nociceptive activation alone, or in combination with vibration, also evoked an increase in stride speed, both in animals that did not roll and in animals that did roll, after rolling. Combining vibration with nociceptive stimulation in animals that selected fast crawl significantly increased stride speed compared to either stimulation alone. Vibration combined with nociceptive stimulation therefore affects both aspects of the escapes response, facilitating the triggering of rolling (Fig. 1e, f) and increasing crawling stride speed in animals that do not roll. Wilcoxon rank sum test with Holm–Bonferroni correction. $**P < 0.0001$. **g**, The synergistic effect on rolling of combining vibration with nociceptive stimulation is observed across a range of frequencies. Rolling probabilities of *MD IV-GAL4>UAS-dTrpA1* (nociceptive and combined conditions) or *MD IV-GAL4>+* (no stimuli and mechanosensory conditions) animals during vibration stimulation at different frequencies, either alone (green) or in combination with thermogenetic activation of nociceptive neurons (32 °C) (grey). See Supplementary Table 3 for n (animals) in this and all other Extended Data Figures. See Supplementary Table 2 for a list of genotypes used all Figures.

a R72F11(Basin) > GFP / MD IV > tdTomato — Basin–MD IV overlap

b R72F11(Basin) > GFP / Ch > tdTomato — Basin–mechano-ch overlap

c Basin-1 | Basin-2 | Basin-3 | Basin-4

d

	Basin-1	Basin-2	Basin-3	Basin-4
Lateral soma location	+	+	+	+
Primary neurite following intersegmental nerve	+	+	+	+
Medial ipsilateral axon	+	+	+	+
Dendrites on ventral side of Ch domain	+	+	+	+
Dendrites on dorsal side of Ch domain	+			
No ventral branch off primary neurite		+		
Ventral dendrites stop lateral to axon position	+	+		
Ventral dendrites extend to or across midline				+
Relative position of axon is ventral	+	+		
Relative position of axon is dorsal			+	+
Presynaptic sites exist on dendritic arbor	+			+

e

	Ch	MD IV
Central lateral terminals	+	
Five cells in intersegmental nerve, three in segmental nerve	+	
Extreme ventromedial terminals		+
One cell in intersegmental nerve, two in segmental nerve		+

f abd1.5 data | L1 data — Interseg. Nerve, Seg. Nerve

g abd1.5 data — 40 μm, 47 μm; L1 data

h presynaptic site / postsynaptic site — abd1.5 data (Left Cell, Right Cell (mirrored)); L1 data (Left Cell, Right Cell (mirrored))

i Anterior ← Segment of Basin → Posterior

	Segment M-1		Segment M	
	MD IV	Ch	MD IV	Ch
Basin-1L	0 (0.00)	40 (0.08)	1 (0.00)	63 (0.13)
Basin-1R	0 (0.00)	31 (0.07)	4 (0.01)	50 (0.12)
Basin-2L	9 (0.03)	22 (0.08)	21 (0.07)	20 (0.07)
Basin-2R	10 (0.03)	15 (0.05)	20 (0.06)	29 (0.09)
Basin-3L	2 (0.01)	11 (0.05)	1 (0.00)	33 (0.14)
Basin-3R	0 (0.00)	18 (0.09)	0 (0.00)	19 (0.10)
Basin-4L	1 (0.00)	3 (0.01)	46 (0.20)	28 (0.12)
Basin-4R	0 (0.00)	5 (0.03)	19 (0.10)	35 (0.18)

abd1.5 data

syn (fraction total input)

	MD IV	Ch	Segment M+1	
			MD IV	Ch
Basin-1L	2 (0.01)	29 (0.08)	0 (0.00)	0 (0.00)
Basin-1R	1 (0.00)	49 (0.13)	1 (0.00)	0 (0.00)
Basin-2L	18 (0.06)	8 (0.03)	1 (0.00)	0 (0.00)
Basin-2R	20 (0.06)	17 (0.05)	0 (0.00)	1 (0.00)
Basin-3L	1 (0.01)	28 (0.17)	0 (0.00)	0 (0.00)
Basin-3R	0 (0.00)	36 (0.15)	0 (0.00)	0 (0.00)
Basin-4L	36 (0.17)	21 (0.10)	7 (0.03)	0 (0.00)
Basin-4R	40 (0.20)	6 (0.03)	7 (0.04)	0 (0.00)

L1 data

**j** Anterior Segment | Same Segment | Posterior Segment — Ipsilateral Side / Contralateral Side — Ch, Basins, MD IV

Extended Data Figure 2 | This figure relates to Fig. 2a, b. a, b, Confocal images of basins (green) and nociceptors (**a**) or chordotonals (**b**) (both in magenta), with cartoons to demonstrate their position within the nerve cord. **c**, Confocal images of individual Basins (in addition to those shown in Fig. 3a) visualized with a FLP-based method. Outline, neuropil boundary. Scale bar, 30 µm. **d**, Table of morphological features based on which Basins are uniquely distinguishable from other interneurons (blue) and individual Basin types from each other (purple). **e**, The table of morphological features based on which the mechanosensory chordotonal (Ch; green) and the nociceptive (orange) neuron projections are uniquely identifiable. **f**, Electron microscopy reconstructions of MD IV (orange) and chordotonal (green) axons in two abdominal segments, from two different animals. Top, reconstructions from the A3 segment in the smaller electron microscopy volume (comprising 1.5 abdominal segments: A3 and a part of A2) shown in Fig. 2. Bottom, reconstructions from the segment A1 in the comprehensive electron microscopy volume that spans the entire nervous system (L1 electron microscopy volume, see Fig. 5 and Supplementary Video 7). Outlines indicate approximate neuropil boundaries. Scale bar, 20 µm. **g**, Electron microscopy reconstructions of nociceptive (orange) and mechanosensory (green) axons and of the four Basin types (black) in the two abdominal segments, from two different animals. Top, A3 segment from the smaller 1.5 abdominal segment volume. Bottom, A1 segment in the comprehensive L1 electron microscopy volume. Outlines indicate approximate neuropil boundaries. Scale bar, 20 µm. The MD IV and chordotonal terminals span the ventromedial and ventrocentral nerve cord domains, respectively. Basin-2 and Basin-4 dendrites overlap with chordotonal and MD IV terminals. **h**, Detailed views of the neuronal arbors and synaptic locations of each Basin cell type from the two segments (from the two animals). Cells on the right side are mirrored for easier comparison. The location of synaptic domains is conserved across hemisegmental homologues and animals. The dendritic arbors of individual Basins across all four hemisegments are highly stereotyped and uniquely recognizable. **i**, Tables show the numbers of synapses that Basins in the A3 segment (from the smaller 1.5 segment electron microscopy dataset, top) and in the A1 segment (from the comprehensive L1 electron microscopy volume, bottom) receive from MD IV and chordotonal neurons in their own and in adjacent abdominal segments. See also Supplementary Tables 4a and 5a for more information. Both the left and right Basin-1 and Basin-3 homologues reproducibly received many synapses (each >25 synapses and >15% of total input, on average) from chordotonal in both animals, but very few (no more than 1% of total input synapses) from MD IV neurons. Both the left and right Basin-2 and Basin-4 homologues received many inputs from both chordotonal (on average >20 synapses and >10% total input) and MD IV (on average >20 synapses and >10% total input) neurons. **j**, Schematic representation of the distribution of inputs from MD IV and chordotonal from adjacent segments onto Basins.

a

b

	Total inputs
A02m abd1.5 L	69
A02m abd1.5 R	55
A02m L1 L	50
A02m L1 R	73
A02n abd1.5 L	81
A02n abd1.5 R	55
A02n L1 L	74
A02n L1 R	80
A10a abd1.5 L	162
A10a abd1.5 R	178
A10a L1 L	103
A10a L1 R	105

c

A02m+A02n

	Ch	MD IV	Basin-2	Basin-4
abd1.5 L	6	31	15	0
abd1.5 R	5	18	6	0
L1 L	3	29	15	0
L1 R	6	15	24	0

A10a

	Ch	MD IV	Basin-2	Basin-4
abd1.5 L	0	41	0	2
abd1.5 R	0	32	0	3
L1 L	0	20	1	2
L1 R	0	26	0.5	3.5

Basin-1

	Ch	MD IV	Basin-2	Basin-4
abd1.5 L	63	1	0.5	0
abd1.5 R	50	4	0	0
L1 L	29	2	0	0
L1 R	49	1	0	0.5

Basin-2

	Ch	MD IV	Basin-2	Basin-4
abd1.5 L	20	21	0	0
abd1.5 R	29	20	0	0
L1 L	8	18	0	0
L1 R	17	20	0	0

Basin-3

	Ch	MD IV	Basin-2	Basin-4
abd1.5 L	33	1	0	0
abd1.5 R	19	0	0	1
L1 L	28	1	0	0
L1 R	36	0	0	0

- green: 2+ syn. both hemisegments
- light green: >0 syn. both hemisegments
- yellow: 2+ syn one hemisegment, other animal 2+ syn.

d

A02m/n

	vchA/B (2)	v'ch	lch5-1	lch5-3	lch5-5	lch5-2/4 (2)	vdaB	v'ada	ddaC	Basin-2	Basin-4
abd1.5 L	0	3	0	0	0	0	3	1	14	6	0
abd1.5 R	0	1	1	0	0	0	0	1	12	1	0
L1 L	0	0	0	0	0	0	3	3	5	10	0
L1 R	0	4	0	0	0	0	0	0	3	10	0

A02n/m

	vchA/B (2)	v'ch	lch5-1	lch5-3	lch5-5	lch5-2/4 (2)	vdaB	v'ada	ddaC	Basin-2	Basin-4
abd1.5 L	0	3	0	0	0	0	1	3	9	9	0
abd1.5 R	0	3	0	0	0	0	1	4	5	0	0
L1 L	2	1	0	0	0	0	3	7	8	5	0
L1 R	0	2	0	0	0	0	1	1	10	14	0

Basin-1

	vchA/B (2)	v'ch	lch5-1	lch5-3	lch5-5	lch5-2/4 (2)	vdaB	v'ada	ddaC	Basin-2	Basin-4
abd1.5 L	10	18	12	7	1	15	1	0	0	1	0
abd1.5 R	9	13	11	9	1	7	0	3	1	0	0
L1 L	6	11	4	1	0	7	2	0	0	0	0
L1 R	10	11	9	5	0	14	1	0	0	0	1

Basin-2

	vchA/B (2)	v'ch	lch5-1	lch5-3	lch5-5	lch5-2/4 (2)	vdaB	v'ada	ddaC	Basin-2	Basin-4
abd1.5 L	2	13	4	1	0	0	3	11	7	0	0
abd1.5 R	4	13	3	9	0	0	10	0	10	0	0
L1 L	1	2	3	2	0	0	2	12	4	0	0
L1 R	0	13	1	0	1	2	1	11	8	0	0

Basin-3

|	vchA/B (2)	v'ch	lch5-1	lch5-3	lch5-5	lch5-2/4 (2)	vdaB	v'ada	ddaC	Basin-2	Basin-4
abd1.5 L	20	6	6	1	0	0	1	0	0	0	0
abd1.5 R	6	4	6	2	0	1	0	0	0	0	2
L1 L	21	0	4	3	0	0	1	0	0	0	0
L1 R	20	3	13	0	0	0	0	0	0	0	0

Extended Data Figure 3 | Reproducibility of the wiring diagram across animals. a, Electron microscopy reconstructions of the local interneurons downstream of both the MD IV and Basin neurons in both electron microscopy volumes; the volume comprising 1.5 abdominal (A2/A3) nerve cord segments of one animal (abd1.5, Fig. 2e), and the electron microscopy volume comprising the entire nervous system of another (L1, Fig. 5a and Supplementary Video 7). These local neurons (as well as the Basins shown in Fig. 2f and Extended Data Fig. 2h) with cell bodies and the majority of their arbor in the same segment as the sensory neurons are readily identifiable in both volumes and can be used to assess the reproducibility of sensory–interneuron and interneuron–interneuron connectivity across animals. Three distinct short-range interneuron types (A02m, A02n and A10a) received at least two synapses from nociceptive sensory neurons in both animals. A02m/n and A10a neurons also received inputs from Basin-2 or Basin-4 (Supplementary Information (Supplementary Atlas section)). The two interneurons A02m and A02n belong to the same lineage (lineage 2), have indistinguishable projections and connectivity, and can only be distinguished based on a small difference in the dorsoventral cell body location. In contrast, Basin-1, Basin-2, Basin-3 and Basin-4, even though they belong to the same lineage (lineage 9), all have distinct projections (and connectivity) and can each be considered a unique

neuron type. A10a also have unique projections and belong to lineage 10. **b,** Total number of inputs each neuron in **a** receives. **c,** Reproducibility of connections between neuron types across individuals. Table shows class level connectivity between neurons in the abd1.5 and L1 data sets. Values given are the total sum of all synapses for all members of each class. 100% of connections with at least two synapses between homologous neuron types that were reproducible across the same individual were also reproducible across individuals; that is, the connection between those neurons existed in the other individual. **d,** Reproducibility of connections between individual neurons across individuals. Table shows connectivity between individual neurons in the abd1.5 and L1 data sets. Whenever two or more synapses were found between homologous neurons in both hemisegments in one animal, either two or more were also found in both hemisegments in the other animal (dark green cells); or at least one synapse per edge was found in both hemisegments in the other animal (light green cells); or two or more synapses were found in at least one hemisegment of the other animal (yellow cells). Thus, in 100% of cases in which a connection was present between homologues on both sides in one individual it was also present between homologues in the other individual, at least on one side.
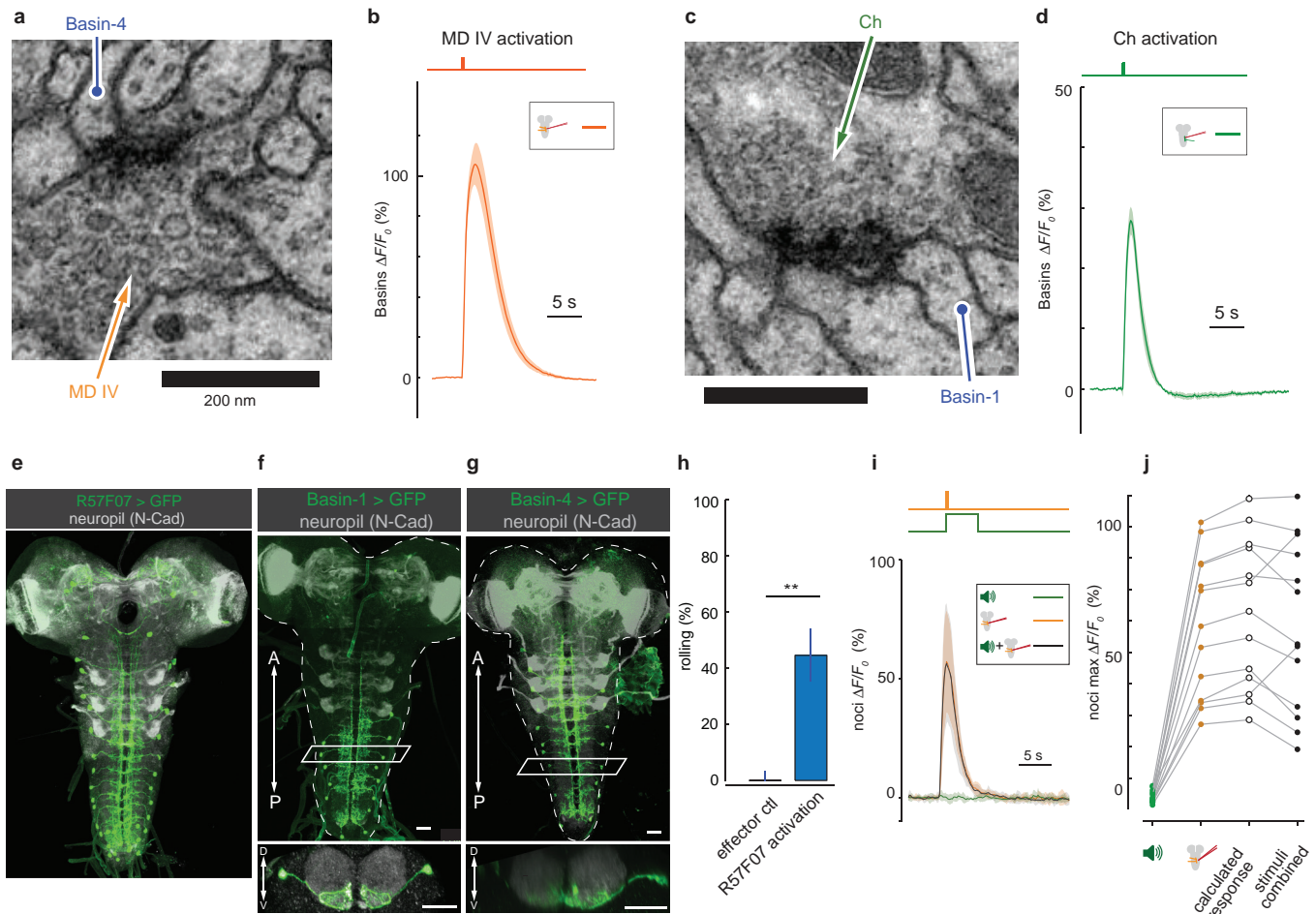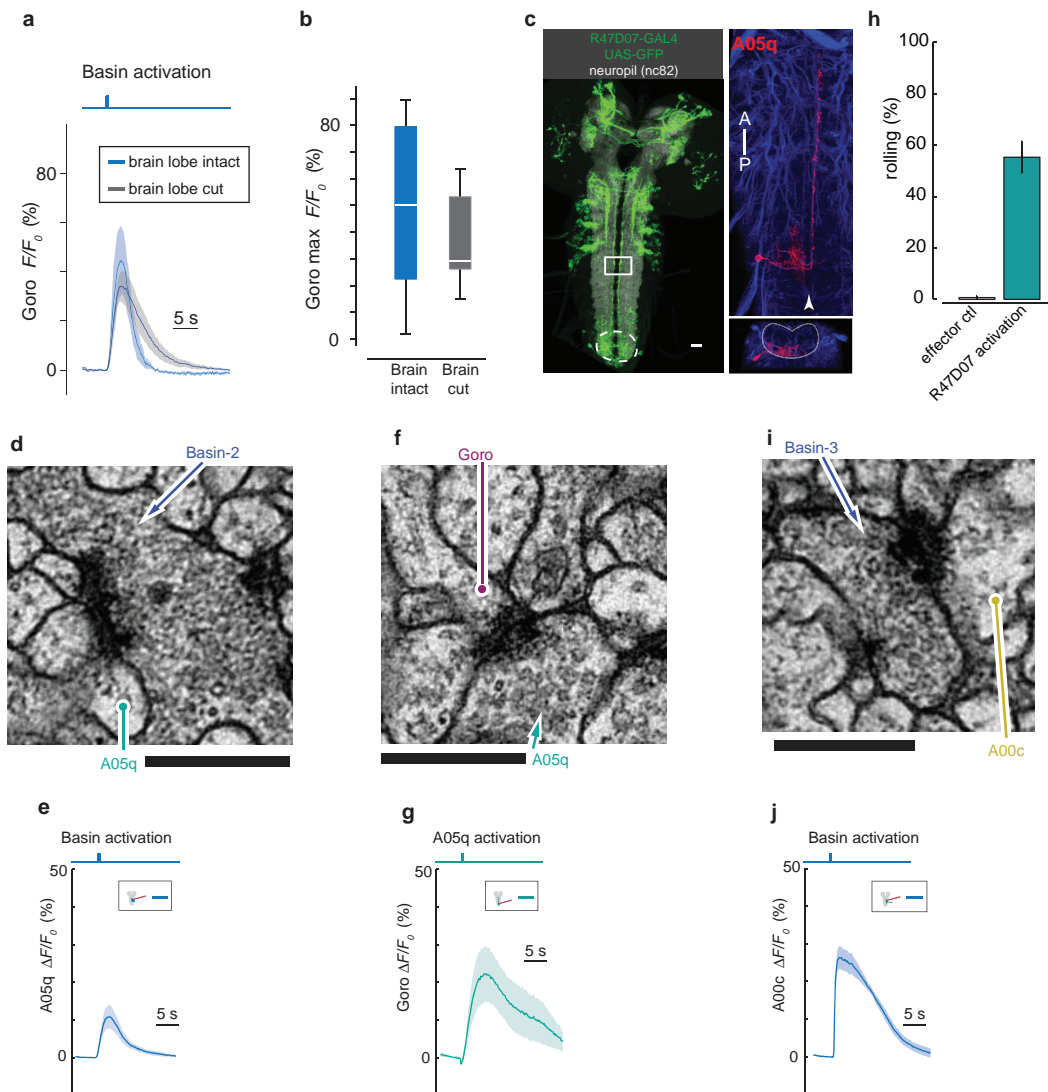
Extended Data Figure 4 | This figure relates to Figs 2 and 3. a, c, Electron micrographs of representative synapses formed by MD IV and chordotonal (Ch) projections onto Basins containing a T-bar ribbon. Circles point to the postsynaptic neuron and arrows point to vesicle clouds in the presynaptic terminals. All synapses have small clear-core vesicles associated with fast chemical neurotransmitters such as acetylcholine, glutamate, or GABA. Scale bars are 200 nm. **b, d,** Calcium transients evoked in Basins by MD IV (**b**) and chordotonal (**d**) activation by local ATP injection. The ATP-gated channel, P2X₂, is expressed in MD IV (**b**) and chordotonal (**d**) neurons and GCaMP6 is expressed in Basins (in *MD IV-lexA>lexAop-P2X₂; R72F11-GAL4>UAS-GCaMP6* (**b**) and *Ch-lexA>lexAop-P2X₂; R72F11-GAL4>UAS-GCaMP6* (**d**) animals). $n = 30$ trials (**b**), and 18 trials (**d**). Functional connectivity experiments suggest the connection from sensory neurons onto Basins is excitatory. Since chordotonal neurons are the only neurons that respond to vibration[32,35] we used vibration in subsequent experiments so that we could activate chordotonal with vibration, and MD IV with P2X₂ and ATP injection,

in the same animal. **e,** *R57F07* drives expression in Basin-4 neurons and in a few brain neurons. *R57F07* was therefore used for generating the Basin-4 selective split-GAL4 line shown in **g. g, h,** Expression patterns of driver lines selective for Basin-1 (*R20B01*) and Basin-4 (*R72F11* and *R57F07* intersection) neurons (green). Scale bar, 25 μm. **f,** Activation of *R57F07* is sufficient to evoke rolling in 45% of animals. Intersecting the parent line *R57F07* with *R72F11* (both drive expression in Basin-4 neurons and evoke rolling) results in a split-GAL4 line that drives expression selectively in Basin-4 neurons, shown in **g. i, j,** Negative control for Fig. 4d, f. **i,** Calcium transients in MD IV terminals observed in response to vibration (1,000 Hz, 6.8 m s⁻²), MD IV activation by local ATP injection or their combination (in *MD IV-lexA>lexAop-P2X₂; MD IV-GAL4>UAS-GCaMP6* animals). $n = 14$ trials. **j,** Maximum $\Delta F/F$ of calcium transients in response to mechanosensory, nociceptive, and combined simulation from **i**. Unlike Basin-4 (Fig. 3b, c), the MD IV neurons do not respond more to the combination of stimuli compared to MD IV activation alone.
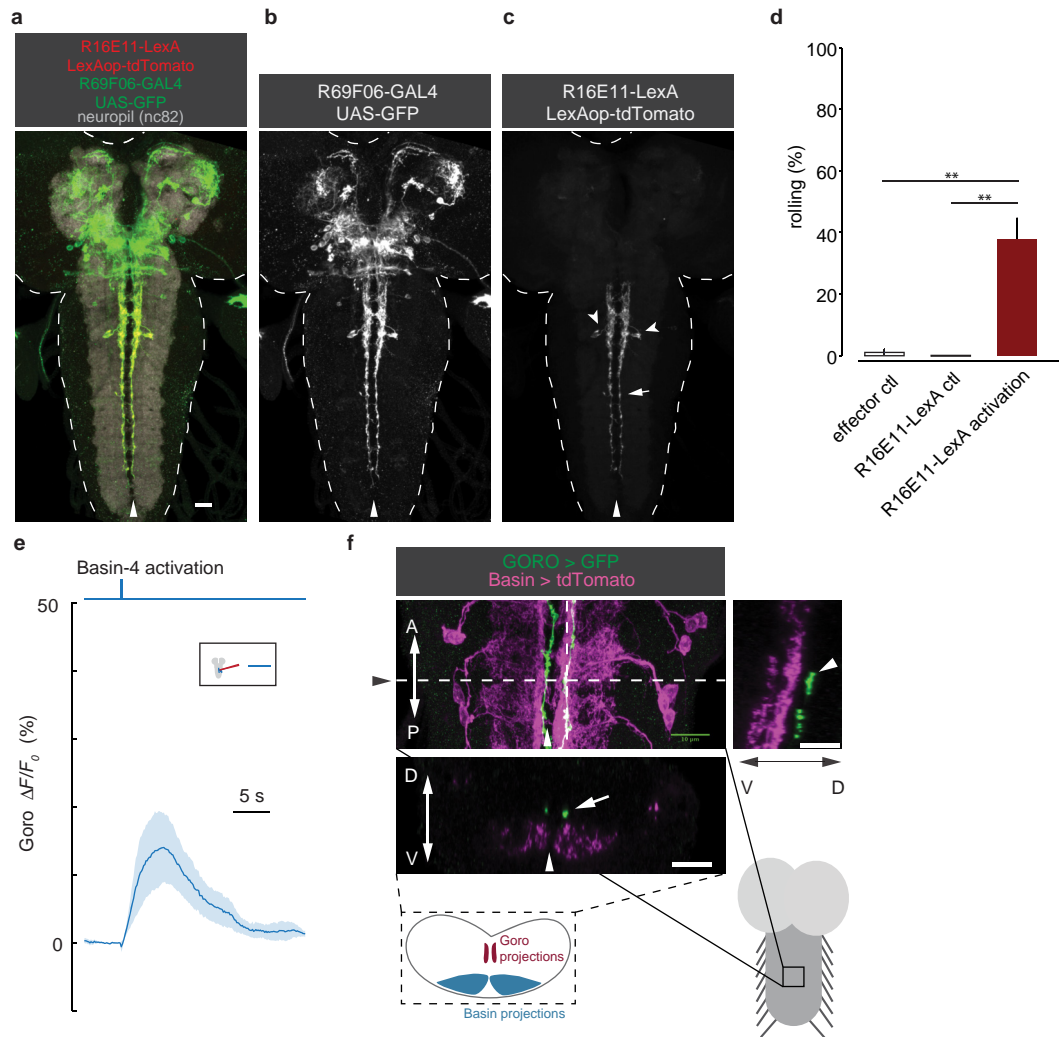
Extended Data Figure 5 | This figure relates to Fig. 4. a–c, The *R16E11-LexA* line faintly, but selectively, drives expression in the same pair of thoracic neurons (Goro neurons) as the *R69F06 GAL4* line. **a**, Co-localization of *R16E11-LexA* and *R69F06-GAL4* expression in the same nervous system (genotype: *R16E11-lexA/LexAop-mry-tdTomato;R69F06-GAL4,UAS-mry-GFP*). *R16E11-lexA>lexAop-tdTomato* (red) and *R69F06-GAL4>UAS-GFP* (green) co-localize in the two Goro cell bodies and their projections in the nerve cord (yellow). **b, c**, Single-channel images from **a**. Scale bar, 20 μm. **d**, Thermogenetic activation (32 °C) of the single pair of Goro neurons using the *R16E11* line (*R16E11-lexA>lexop-dTRPA1*) evoked rolling in 36% of animals. Activating neurons in *R69F06* evokes even more rolling (86%, Fig. 4a). *R69F06* drives expression in Goro neurons more strongly that *R16E11*, but it

also drives expression in more neurons. Using a FLP-based intersection between *R16E11-lexA* and *R69F06* (see Methods), we found that 76% of animals with expression of dTRPA1 selectively in Goro rolled when exposed to restrictive temperature (32 °C). Thus, activating the single pair of command-like Goro neurons evokes rolling. **e**, Calcium transients in Goro dendrites in response to selective Basin-4 activation by local ATP injection ($n = 12$ trials). The ATP-gated $P2X_2$ channel and GCaMP6 are expressed selectively in Basin-4 and Goro neurons, respectively (in *Basin-4-Gal4>UAS-P2X_2; Goro-lexA(R69F06-lexA)>lexAop-GCaMP6* animals). **f**, Basin (magenta) and Goro (green) projections do not co-localize (in *16E11-lexA>LexAop-tdTomato; R72F11-GAL4>UAS-GFP* larvae), indicating that the two cell types that are functionally connected cannot be directly synaptically connected.

Extended Data Figure 6 | This figure relates to Fig. 5. a, The nerve cord Basin–Goro pathway is sufficient to activate the Goro neurons. Calcium transients in Goro dendrites evoked by Basin activation by local ATP injection with the brain lobes intact ($n = 15$ trials) or removed ($n = 24$ trials) (in *R72F11-LexA>LexAop-P2X2; Goro-GAL4(R69F06)>GCaMP6* animals). **b,** Quantification of **a**. The difference is not significant. Error bars show s.e.m. **c,** Left, *R47D07* drives expression in A05q neurons (that receive direct inputs from Basin-2 and synapse onto Goro neurons) and in other neurons in thorax and brain (green, visualized with anti-GFP). In abdominal segments A8 and A7, the only neurons labelled with this line are the A05q neurons (dashed circle). The rest of the nervous system (grey) is visualized with anti-N-cadherin. White square, A05q axon terminals that synapse onto Goro dendrites. Scale bar, 25 μm. Note that this line also drives expression in developing adult neurons in thorax and brain that are not functional in the larval nervous system (secondary lineages). Right, image of individual A05q cells (magenta) visualized with a FLP-based strategy in *R47D07*, in top-down (top) and cross-section (bottom) views. Arrowheads indicate midline. Thin white lines indicate neuropil boundary. **d, f, i,** Electron micrographs of representative synapses formed by

Basin-2 onto A05q (**d**), A05q onto Goro (**f**) and Basins onto A00c (**i**) containing T-bar ribbons. Circles point to the postsynaptic neuron and arrows point to vesicle clouds in the presynaptic terminals. All synapses have small clear-core vesicles associated with fast chemical neurotransmitters such as acetylcholine, glutamate, or GABA. Scale bars, 200 nm. **e,** Calcium transients in A05q axon terminals (square in **c**) evoked by Basin activation using local ATP injection (*R72F11-LexA>lexop-P2X2; R47D07-GAL4>UAS-GCaMP6, n = 18* trials) suggest a functionally excitatory connection. **g,** Calcium transients in Goro neurons evoked by A05q activation using local ATP injection specifically in the vicinity of A05q dendrites in A8 (dotted circle in **c**; *R47D07-GAL4>UAS-P2X2; Goro(R69E06-LexA)>lexop-GCaMP6, n = 18* trials) suggest a functionally excitatory connection. **h,** Thermogenetic activation of neurons in *R47D07* triggers rolling (exposing *R47D07-GAL4>UAS-dTrpA1* animals to 32 °C). Control, *+>UAS-dTrpA1*. **j,** Calcium transients in A00c evoked by Basin activation using local ATP injection (*R72F11-LexA>lexop-P2X2; R71A10-GAL4>UAS-GCaMP6, n = 15* trials) suggest a functionally excitatory connection.
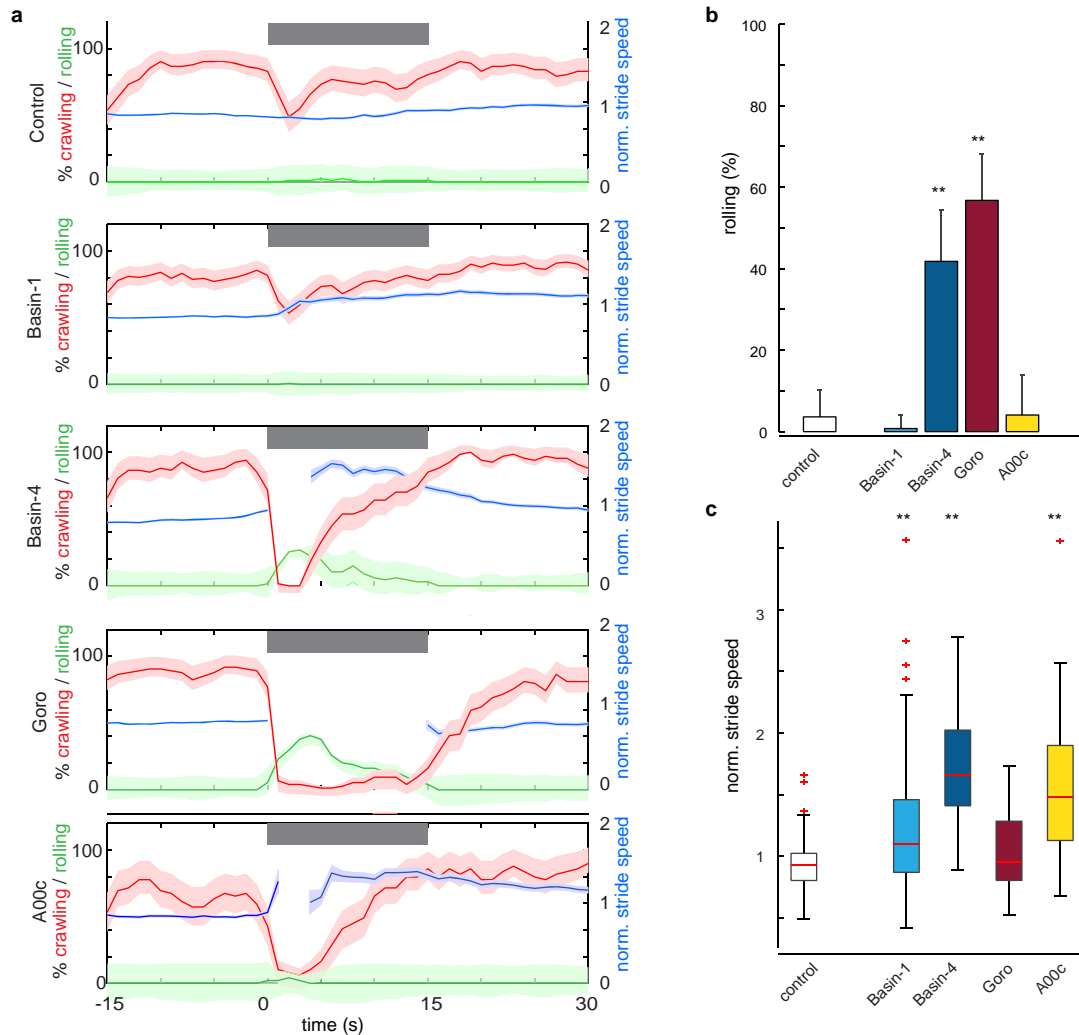
Extended Data Figure 7 | Distinct neurons differentially control distinct aspects of the escape sequence. a, Graphs show the time course of rolling (green) and crawling (red and blue) behaviours before, during and after a 15 sec (indicated by the grey box) optogenetic (617 nm red light) activation period. The percentage of larvae that are rolling (green line) or crawling (red line) at each time point is shown. The normalized speed of crawling strides (normalized by speed prior to optogenetic stimulation) is also shown for those animals that are crawling at the time (blue line). Means (dark line) and s.e.m. (shaded areas) are shown. Data are from $+>UAS\text{-}CsChrimson$, $basin1\text{-}GAL4>UAS\text{-}CsChrimson$, $basin4\text{-}GAL4>UAS\text{-}CsChrimson$, $Goro\text{-}GAL4(R69F06)>UAS\text{-}CsChrimson$, or $A00c\text{-}GAL4(R71A10)>UAS\text{-}CsChrimson$ animals, from top to bottom. **b**, Percentage of larvae that rolled at least once during the entire 15 sec activation window. Error bars are 95% confidence intervals. $**P < 0.0001$. Optogenetic activation of Basin-4 and Goro neurons evokes rolling, consistent with the results of thermogentic activation experiments shown in Figs 2a, g and 4e. Note that the instantaneous percentage of rolling shown in **a** is always lower than the percentage shown in **b** because not all larvae roll at exactly the same time. **c**, Comparison of the normalized speed of crawling strides in response to optogenetic activation. Boxes show medians (red) with 25th and 75th quantiles of the data. Whisker is 1.5 interquartile range (IQR) for the length of the whiskers. Red plus is outlier. $**P < 0.0001$. We observe that in control larvae, red light alone evoked a mild reduction in crawling speed, whereas Basin-1 activation evoked a mild, but significant, increase in stride speed and did not evoke rolling. Basin-4 activation evoked the entire escape sequence (roll followed by fast crawl). Note first an increase in instantaneous percentage of rolling larvae (peak in green line in **a**), followed by an increase in the instantaneous percentage of crawling larvae, and an increase in stride speed relative to the period prior to stimulation (peak in blue line in **a**). Goro neuron activation only evoked rolling and not crawling (note the percentage of crawling larvae is almost 0 during the optogenetic activation period). Furthermore there was no increase in relative stride speed following stimulation offset. Activation of the three A00c neurons with $R71A10$ was not sufficient to evoke rolling, but it was sufficient to evoke a mild, but significant, increase in stride speed. Thus, distinct neurons downstream of Basins appear to differentially control distinct aspects of the 'roll-fast crawl' escape sequence.
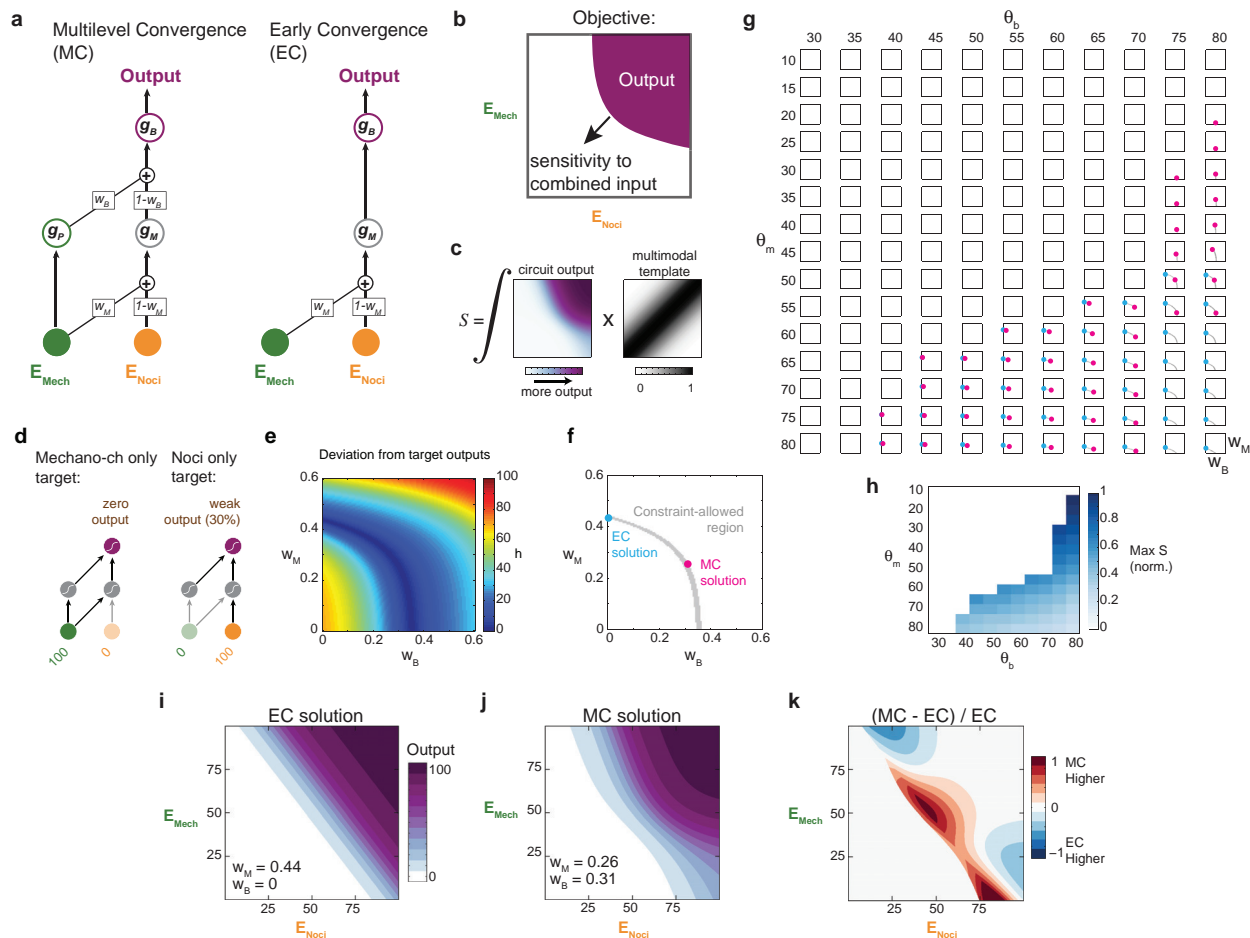
Extended Data Figure 8 | Multilevel convergence can enhance sensitivity to weak multimodal stimuli. A benefit of the multilevel architecture could be to amplify the effect of mechanosensory cues on nociceptive cues, increasing the sensitivity to relatively weak bimodal cues. To explore this idea further, we used a simple model to ask whether a two-level network with two levels of convergence (multilevel convergence; MC) can be more sensitive to relatively weak bimodal events than a network with only early convergence (EC).
a, Schematic for a simple model of early and multilevel convergence networks (see Methods for details). We considered an early convergence and a multilevel convergence two-layer feed-forward circuit with two inputs, corresponding to MD IV (E_{Noci}) and chordotonal (E_{Mech}), and one output corresponding to rolling occurrence. We modelled steady-state firing rates, where the output of each model neuron is a sigmoidal logistic function (with a lower threshold and upper saturation) of a weighted sum of its inputs. One pathway remains only chordotonal, while the other mixes modalities only early, or at both levels, depending on weights (w_M and w_B). **b–d,** Solutions for w_M and w_B are found using a constrained optimization procedure that maximizes sensitivity to weak bimodal inputs (**b, c**) within two experimentally observed unimodal target outputs (**d**): (1) that mechanosensory stimulus alone never evokes rolling; and (2) that the strongest MD IV stimulus alone evokes only 30% output (Fig. 1e, f and see Methods for details). **e,** Example deviation h from target outputs (see Methods) for multiple values of w_M and w_B for one set of thresholds ($\theta_P = 40$; $\theta_M = 50$; $\theta_B = 75$). **f,** Values of w_M and w_B that maximize sensitivity S (dots) while also satisfying $h < 3$ (grey area). **g,** Early and multilevel solutions (if they exist) for other thresholds. Solutions exist if one or both θ_M or θ_B is high

(keeping $\theta_P = 40$). Note that although the optimal sensitivity multilevel convergence solution could have $w_M = 0$ and thus effectively be an early convergence solution, this does not occur; multilevel convergence solutions are always more sensitive when they exist. Multilevel convergence solutions with $w_B = 0$ are not shown since they do not exhibit multilevel convergence.
h, Sensitivity of the optimal multilevel convergence solution as a function of the same thresholds as in Extended Data Fig. 8g, normalized by the most sensitive multilevel convergence solution found. Multilevel convergence solutions are more sensitive to relatively weak multimodal stimuli than early convergence solutions. Across all thresholds tested, the most sensitive circuits occur for $\theta_M = 50$ and $\theta_B = 75$ (Extended Data Fig. 8k). For such parameters, no early convergence solutions satisfy the unimodal constraints, hence multilevel convergence solutions are overall the most sensitive. **i, j,** Example early convergence and multilevel convergence solutions for w_B and w_B that satisfy the condition in **d** for one set of neuronal firing thresholds ($\theta_P = 40$; $\theta_M = 50$; $\theta_B = 75$). **k,** Subtracting the output of early convergence from the multilevel convergence circuit. The multilevel convergence circuit triggers more rolling than the early convergence circuit in response to relatively weak bimodal cues and strong unimodal MD IV cues, but less rolling than the early convergence circuit to strong unimodal chordotonal cues (top left corner). The multilevel convergence circuit thus offers a more complex response function with greater sensitivity and selectivity for multimodal and strong unimodal MD IV cues, but not chordotonal cues, and could enable enhanced selection of rolling to more threatening events and reduced selection of rolling to less threatening events.

ARTICLE

Structure of the human 80S ribosome

Heena Khatter[1,2,3,4]*, Alexander G. Myasnikov[1,2,3,4]*, S. Kundhavai Natchiar[1,2,3,4]* & Bruno P. Klaholz[1,2,3,4]

Ribosomes are translational machineries that catalyse protein synthesis. Ribosome structures from various species are known at the atomic level, but obtaining the structure of the human ribosome has remained a challenge; efforts to address this would be highly relevant with regard to human diseases. Here we report the near-atomic structure of the human ribosome derived from high-resolution single-particle cryo-electron microscopy and atomic model building. The structure has an average resolution of 3.6 Å, reaching 2.9 Å resolution in the most stable regions. It provides unprecedented insights into ribosomal RNA entities and amino acid side chains, notably of the transfer RNA binding sites and specific molecular interactions with the exit site tRNA. It reveals atomic details of the subunit interface, which is seen to remodel strongly upon rotational movements of the ribosomal subunits. Furthermore, the structure paves the way for analysing antibiotic side effects and diseases associated with deregulated protein synthesis.

Ribosomes form the hub of translation across all kingdoms of life, synthesizing proteins using an mRNA template. An interesting concept that has emerged from previous ribosome structures is its layered evolution, with a conserved core across all species comprising the catalytic peptidyl transferase centre (PTC) and GTPase- (guanosine triphosphatase) associated centre in the large ribosomal subunit, and the decoding centre in the small subunit[1]. The exceptionally complex ribosomes of multicellular eukaryotes contain additional rRNA components, termed expansion segments (ES), and additional protein moieties as compared to their prokaryotic counterparts. The human ribosome (80S) has a molecular weight of 4.3 MDa: the large subunit (60S) consists of 28S, 5S and 5.8S rRNAs and 47 proteins, while the small subunit (40S) possesses a single 18S rRNA chain and 33 proteins. The structures of mammalian cytoplasmic ribosomes have provided locations and folds of all eukaryote-specific components[1–7]. A recent structure of the human ribosome at an average resolution of ~5.4–6.0 Å provided useful insights[1]; however, structural data closer to 3.0 Å resolution are required to derive an atomic model of the human ribosome, the ultimate goal of eukaryotic ribosome structural studies. The ribosome is a target for various human diseases (notably infectious diseases and cancer), which gives a structure at or around 3.0 Å resolution added medical importance, especially with regard to the growing problems of antibiotic resistance and drug therapy side effects.

Over the past year, structure determination of asymmetric objects such as the ribosome by single-particle cryo-electron microscopy (cryo-EM) has reached the 3–4 Å resolution range (for example, refs 8–11), owing largely to improved electron detectors[12–14]. Nevertheless, the potential of cryo-EM-associated atomic model building and refinement procedures as used in the field of X-ray crystallography has not been fully exploited, especially in terms of structure refinement of large macromolecular complexes. Here we present the near-atomic structure of the human ribosome determined using high-resolution electron microscopy and a direct electron detector (complementary metal oxide semiconductor (CMOS) detector with a back-thinned direct detection device (DDD) sensor) in combination with recent developments in image processing (dose fractionation and movie processing), 3D reconstruction and atomic model building

using crystallography refinement procedures. The structure of the human 80S ribosome provides unprecedented insights into rRNAs and ribosomal proteins down to the level of individual residue side chains, inter-subunit contacts, functionally relevant key regions including the PTC and the decoding region, interactions with the tRNA present in the exit (E) site, and potential ligand-binding pockets. These can be further exploited for the molecular description of antibiotic and other potential drug pockets.

Structure determination

Human 80S ribosomes from HeLa cells were prepared using a recently established protocol[15] that yields highly homogenous samples of crystallizable quality and which contain a tRNA in the E-site; however, X-ray diffraction of these crystals was limited to ~25 Å (ref. 15). The combination of highly homogenous samples with on-site advanced single-particle cryo-EM imaging techniques (including movie processing done here with only 3 frames) produced a structure at near-atomic resolution (Fig. 1). While the average resolution is estimated to ~3.6 Å (Extended Data Fig. 1), local resolution calculations (Fig. 2) show that most regions are between 2.4 and 4.6 Å resolution, with significant components at 2.9 Å (see features in Extended Data Fig. 2), and only decrease to ~6 Å in the more flexible regions at the periphery. An atomic model of the human ribosome was obtained by extensive model building and restrained parameter refinement protocols[16] (see Methods). We combined this approach with a recent development in the field of X-ray crystallography called feature-enhanced map (FEM[17]), which robustly increases the interpretability of maps. For the first time, FEM map calculations were used in the context of cryo-EM maps for structure interpretation (Fig. 2c, d), which for the well-ordered regions of the human 80S ribosome show resolution and features that go beyond those of the original cryo-EM map. Final electron density maps were calculated by combining phases of the refined atomic model and Fourier coefficients derived from the cryo-EM map, as is routinely done in the crystallography field, yielding further improved maps with features clearly beyond 3 Å resolution (Fig. 2e and Extended Data Figs 3 and 4). This observation indicates that useful amplitudes are present in the cryo-EM data beyond the average resolution estimated by Fourier-shell correlation.

[1]Centre for Integrative Biology (CBI), Department of Integrated Structural Biology, IGBMC (Institute of Genetics and of Molecular and Cellular Biology), 1 rue Laurent Fries, 67404 Illkirch, France. [2]Centre National de la Recherche Scientifique (CNRS), UMR 7104, 67404 Illkirch, France. [3]Institut National de la Santé et de la Recherche Médicale (INSERM) U964, 67404 Illkirch, France. [4]Université de Strasbourg, 67081 Strasbourg, France.
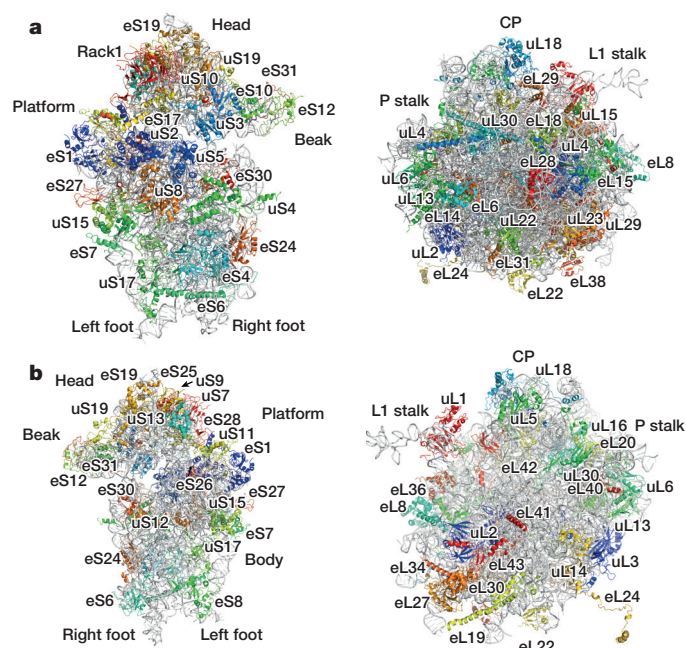*These authors contributed equally to this work.

Figure 1 | Overall structure of the human 80S ribosome. a, b, Global views of the 80S human ribosome from the solvent sides of the 60S and 40S subunits (**a**) and from the inter-subunit interfaces with ribosomal proteins indicated on the atomic model (**b**) (backbone representation; proteins uS14, eS21, eL13, eL21, uL24, eL32, eL33, eL37 and eL39 further inside the structure are not labelled).

The final atomic model (Fig. 1) comprises ∼220,000 atoms across the 5,866 nucleotide residues and ∼11,590 amino acids of the 80 proteins, the 4 rRNAs and E-site tRNA, and 239 Mg^{2+} ions.

Description of the overall structure

The overall structure shows the common landmarks of the 40S and 60S ribosomal subunits (Fig. 1). Compared to the yeast ribosome[2,18], the human ribosome contains an additional protein, eL28 (ref. 1). We observed additional densities corresponding to the carboxy terminus of uS2 (residues 209–222) and also redefined the globular structure of protein eS17 (residues 72–93) in the 40S subunit (Extended Data Fig. 5). The present structure represents a pre-translocation state in which the 40S subunit is non-rotated (Fig. 3); this non-ratcheted state is similar to the yeast ribosome[2], while the recent human and porcine
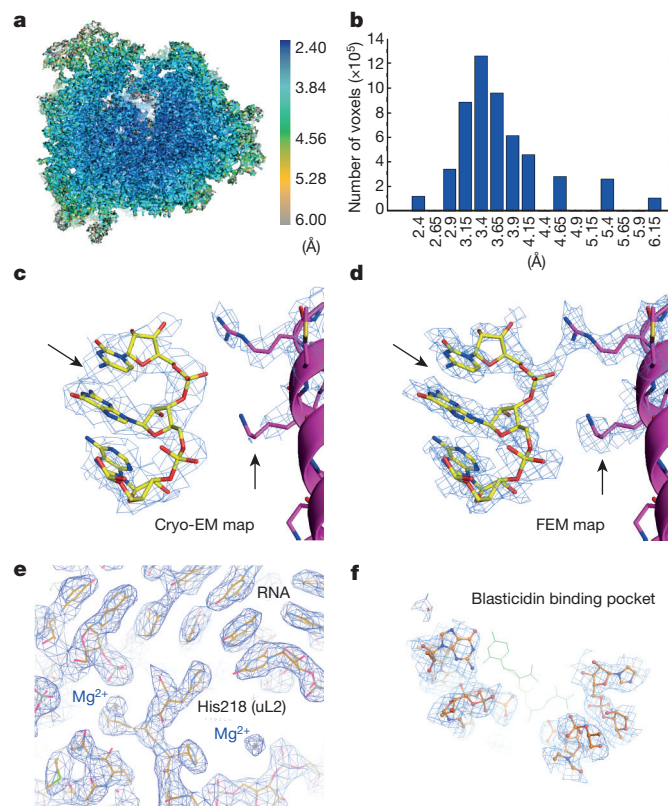
Figure 2 | Local resolution and electron density maps. a, b, Local resolution estimation of the cryo-EM map (central section) and histogram distribution of resolution values. **c, d,** Comparison of cryo-EM (**c**) and FEM (**d**) maps (arrows highlight regions with improved density). **e,** Example of electron density map obtained by combining amplitudes derived from the cryo-EM map and phases calculated from the iteratively refined atomic model (Phenix map calculation), illustrating the interactions of ribosomal protein uL2 with 28S rRNA. **f,** Example of the definition of side-chain residues (Phenix map) forming a ligand-binding pocket (here modelled with blasticidin, a nonspecific antibiotic) as required for structure-guided drug design purposes.

ribosome structures exhibit essentially the same post-translocation conformation due to the presence of elongation factor eEF2 (refs 1, 8). The present conformation is consistent with the presence of a tRNA molecule in the E-site and the movement of the L1 stalk towards the core of the ribosome. The L1 stalk rRNA interacts with the tRNA close to the elbow region, as observed earlier for the 70S
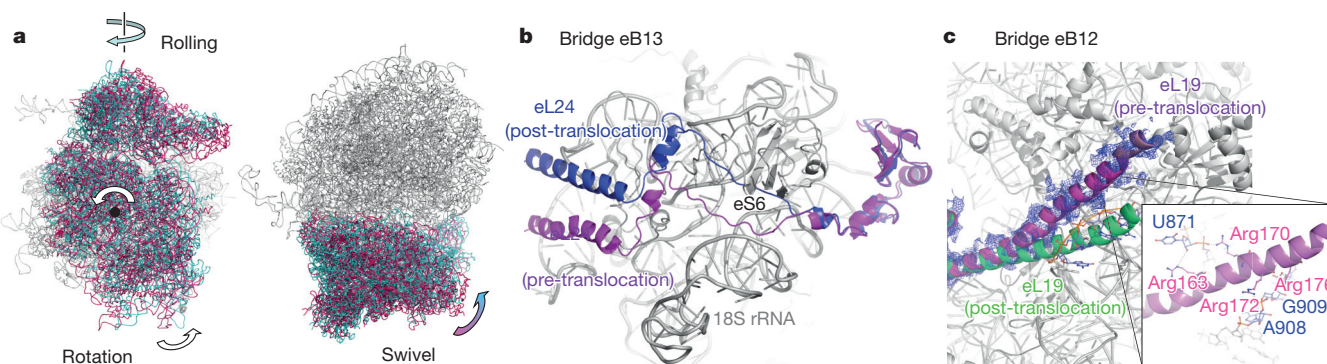

Figure 3 | Conformational changes of the human ribosome. a, Comparison of the human ribosome structure in the pre-translocation state (present study, pink 40S, grey 60S) and in the post-translocation state (when factor eEF2 is present, cyan 40S, porcine ribosome), revealing three major types of movement. **b, c,** Illustration of large conformational changes of protein eL24 (eB13) and

eL19 (eB12) occurring upon 40S subunit rotation. The C-terminal helix of eL24 is relocated to interact with a different region of the 18S rRNA. eL19 undergoes a structural transition from a straight (green, post-translocation state[1,8]) to a bent protein helix (pink, pre-translocation state, present study).

ribosome[19–22]. In contrast to the previous human and porcine ribosome structures, the 40S subunit head shows a swivel movement away from the 60S, which leads to a partial closure of the latch region (Figs 1 and 3a). This could be related with the fact that the mRNA channel is empty, that is, it contains neither mRNA nor protein factors involved in ribosome starvation, in contrast with previous yeast and human ribosome structures[1,2]. In summary, the 40S subunit undergoes three characteristic types of conformational changes: (1) a universal rotation similar to 'ratcheting' in bacterial ribosomes[23]; (2) a swivel movement of the 40S head and beak regions in the direction of the tRNA E-site; and (3) a rotation of the 40S body around a vertical axis perpendicular to the classical rotation axis termed 'subunit rolling' (not observed in bacterial ribosomes, but recently described for a eukaryotic complex[24]; Fig. 3a). The subunit rolling leads to a slight opening of the ribosomal subunit interface (discussed below) on the side of the factor entry site where ribosomal GTPases bind. In this context, it is also interesting to note that relatively large crevasses and gaps exist between the different structural domains of the 18S rRNA, which could serve to accommodate conformational changes of the 40S throughout the translation process.

The transition from a pre- to a post-translocation state induces several major rearrangements in the structure, highlighted by ribosomal proteins eL24 and eL19 (Fig. 3b, c), which constitute some of the characteristic eukaryote-specific bridges (eB12 and eB13, annotated in Fig. 4). The C-terminal part of eL24 relocates entirely and forms a new interaction site on the 40S subunit with ribosomal protein eS6 and 18S rRNA (Fig. 3b). In the case of eL19, a long protein helix protrudes from the 60S subunit to cross over to the 40S subunit. This is one of the rare 60S components (apart from the L1 region) that strongly changes conformation upon subunit rotation. The C-terminal helix of eL19 is kinked in the pre-translocation state (and extends by another seven residues), while it is straight in the post-translocation state (Fig. 3c). This conformational transition results in formation of salt bridges between the positively charged Arg172 and Arg176 side chains of protein eL19 and the phosphate moieties of 18S rRNA nucleotides G909 and G910 (Fig. 3c).

Inter-subunit interface

The present structure highlights some of the characteristic inter-subunit contact sites in more detail with respect to the molecular interactions between the rRNA and protein residues involved. The interface comprises a series of contact regions that have been annotated (Fig. 4, summarized in Extended Data Table 1) according to the nomenclature used for inter-subunit bridges of the yeast ribosome[2] (extended by two new bridges eB15 and eB16[1], not seen in the present structure). The core of the contact interface is formed by helix H69 (28S rRNA, A3766 region; 'h' and 'H' are used to number rRNA helices of the small and large subunits, respectively) from the 60S

Figure 4 | Ribosomal inter-subunit bridges. Annotation of the inter-subunit bridges at the interface of the 40S and 60S subunits (rRNA in light grey, ribosomal proteins in dark grey, colour-coded bridges are indicated).

subunit which interacts with helix h44 (18S rRNA, G1849 region; Fig. 4) to form bridges B2a and B2b, close to the universal rotation axis (Fig. 3a) around which the ribosomal subunits rotate forward and backwards during the tRNA translocation process, as seen in prokaryotic and eukaryotic ribosomes[22,23,25,26]. Among the known inter-subunit bridges, B7b/c and B8 are not observed in the present human 80S ribosome (not including contact-mediating ions such as Os in the yeast 80S structure or Mg^{2+}), while the bridge eB8 contact is mediated by eS1 and eL8, in contrast with the yeast 80S ribosome[2], where eL43 replaces eL8. Furthermore, there is an additional B2e bridge present as observed previously[24]. It is remarkable that a noticeable gap exists between the 40S and 60S subunits in the region below the factor binding site, which is related possibly with the vertical rotation of the 40S subunit and a slight opening of the interface (Fig. 3a). In contrast to other contact regions, there are only few densities observed in the inter-subunit space, which could be ions such as Mg^{2+}. In general, five contact types are observed: ribose–ribose, ribose–phosphate, ribose–base, protein–base and protein–protein interactions (Fig. 5). Below, we highlight some of the most prominent examples of inter-subunit contact regions in the human ribosome.

A particularly interesting contact type are the interactions mediated by ribose moieties, with the 2′-OH groups of the rRNA backbone sugar moieties playing a key role rather than the charged phosphate groups of the rRNA backbone (Fig. 5a). The 2′-OH groups are offset, leading to a bifurcated hydrogen-bonding pattern. This pattern is distinct from sugar-edge (ribose-base) interactions[27] and somehow similar to that of protein β-sheets involving the carbonyl backbone[28–30]. This leads to a characteristic ribose–ribose interface involving mostly hydroxyl groups of the sugar moieties, as illustrated by the interactions of the 2′-OH groups of nucleotides A1719, A1815 and G1816 (18S rRNA helix h44) with the 2′-OH groups of nucleotides G3806, A3807, C3794 and A3795 (28S rRNA helix H71; Fig. 5a). Similar ribose–ribose interactions exist between ribose moieties in the subunit interface of bacterial ribosomes[31–33]. These ribose–ribose contact regions differ from other regions of the interface in which a significant amount of interactions occur through Mg^{2+}-mediated phosphate–phosphate electrostatic interactions, while the bridges located more at the periphery operate through direct hydrogen bonding rather than being mediated by Mg^{2+} (for example, uS13–uL5, Fig. 5, or the eukaryote-specific eL19 and the 18S rRNA 910 region; Fig. 3b, c). This difference could explain the observation that the biochemical stability of eukaryotic ribosomes is much less Mg^{2+}-dependent as compared to bacterial ribosomes, and much more dependent on K^+ ions, consistent with the idea that potassium is required for inter-subunit stability while Mg^{2+} ensures proper rRNA folding[15,34–36]. Characteristic protein–RNA interactions exist, for example, between 18S rRNA region 1173 and the amino-terminal helix of protein eL41 and involve phosphate group interactions with positively charged residues (Arg and Lys; Fig. 5b). Other examples of interface remodelling are regions of the ribosomal subunit interface in which rRNA ribose and Mg^{2+}-mediated contacts sites are also seen to switch positions upon rotation of the ribosomal subunits. This involves rearrangement of the interface and Mg^{2+}-ion repositioning (for example, bridges B2b, 1848 nucleotide region of 18S rRNA, and phosphate backbone of nucleotide 1070; solvent channels with residual densities for more potential ions are found right below the E-site tRNA binding site).

Owing to the conformational difference of the structure as compared to previous human and porcine ribosome structures, numerous contact sites are shifted or completely rearranged. Apart from the two examples of bridges eB12 and eB13 described above that involve major conformational changes of the involved proteins, another prominent example is that of bridge B1b/c (Fig. 5c). This contact is located between the central protuberance of the 60S and the 40S head and involves ribosomal proteins uL5 and uS13 (Fig. 5c). Rather than operating through large conformational changes of parts of the
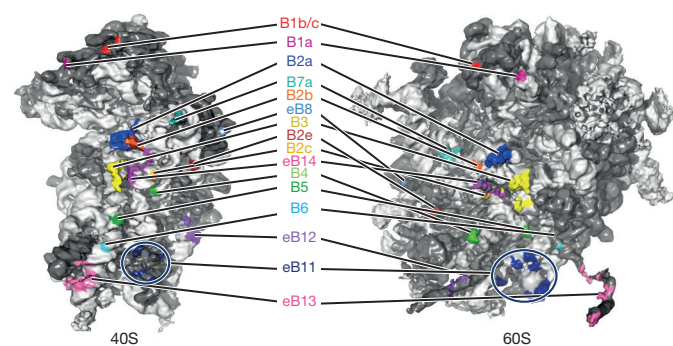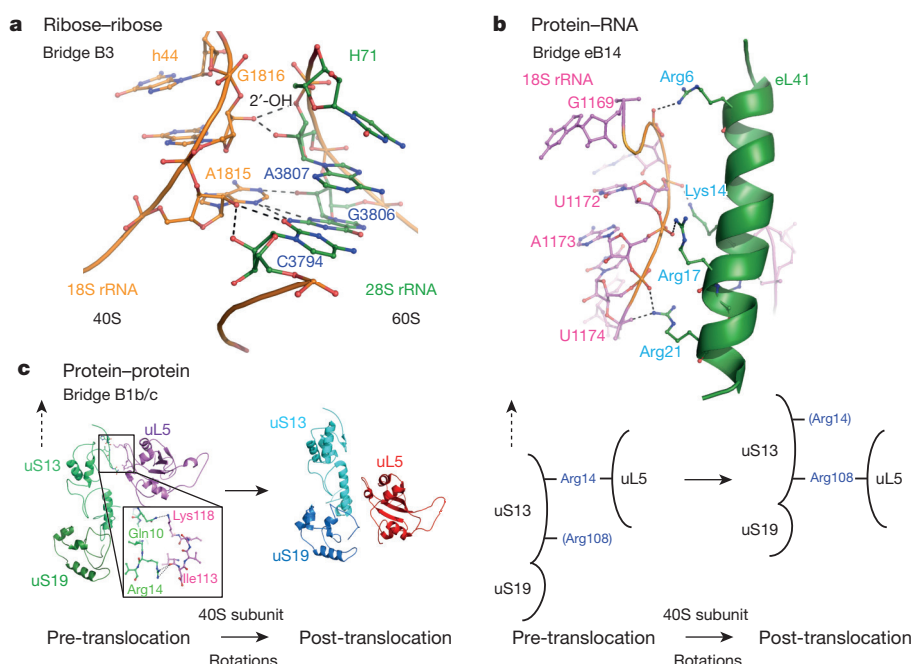
Figure 5 | Molecular details of characteristic inter-subunit contacts.
a, Detail of bridge B3 highlighting ribose–ribose and ribose–nucleotide interactions between 28S rRNA helix H71 and 18S rRNA helix h44, mediated through the 2′-OH ribose moieties. **b**, Characteristic protein–RNA interaction between 18S rRNA region 1173 and the N-terminal helix of protein eL41 (bridge eB14) involving phosphate group interactions with positively charged proteins, this bridge rearranges through a protein shift that is amp-lified at the periphery of the 40S subunit when rotating. Upon subunit rotation and 40S head swivelling the molecular interactions between uS13 and uL5 are completely remodelled and the interacting residues of uS13 shift by about 23–26 Å (Cα atom of Arg14 and Arg104, respectively). In the pre-translocation state, Ile113 and Lys118 (uL5) contact the side chains of Arg14 and Gln10 (uS13). By contrast, in the post-translocation state as observed in the previous human and porcine ribosome structures, this bridge is formed by Tyr119 of uL5 with Arg108 of uS13 (see schematic in Fig. 5c); the large shift allows additional interactions between uL5 and uS19 to be formed. Similar rearrangements have been observed in yeast and *Escherichia coli* ribosomes[2,31,37].

residues. **c**, Details of bridge B1b/c which involves proteins uS13 and uL5 close to the central protuberance of the ribosome. The schematic drawing illustrates the remodelling of the inter-subunit interface that occurs at the atomic level between the pre- and post-translocation states (see also Fig. 3; present study and refs. 1, 8, respectively).

tRNA binding sites and molecular recognition in E-site

The present human 80S ribosome contains a co-purified tRNA in the E-site, which is well-defined in the cryo-EM map and appears to stabilize the L1 stalk (Fig. 6a). Comparison with other E-site tRNA structures shows that its conformation is different in the anti-codon loop regions of the decoding stem, while it is similar at the CCA end of the acceptor stem (Fig. 6b). The tRNA is found in the classical E/E state consistent with the pre-translocation state of the ribosome, while it is observed in a P/E hybrid state in the porcine ribosome[8,38,39] (the presence of tRNA at each of these sites is annotated as E/E or P/E, with the first symbol denoting the contact with the small subunit and the second referring to the large subunit). On the 60S subunit, the E-site tRNA interacts with the L1 stalk oriented inwards, and with the L1
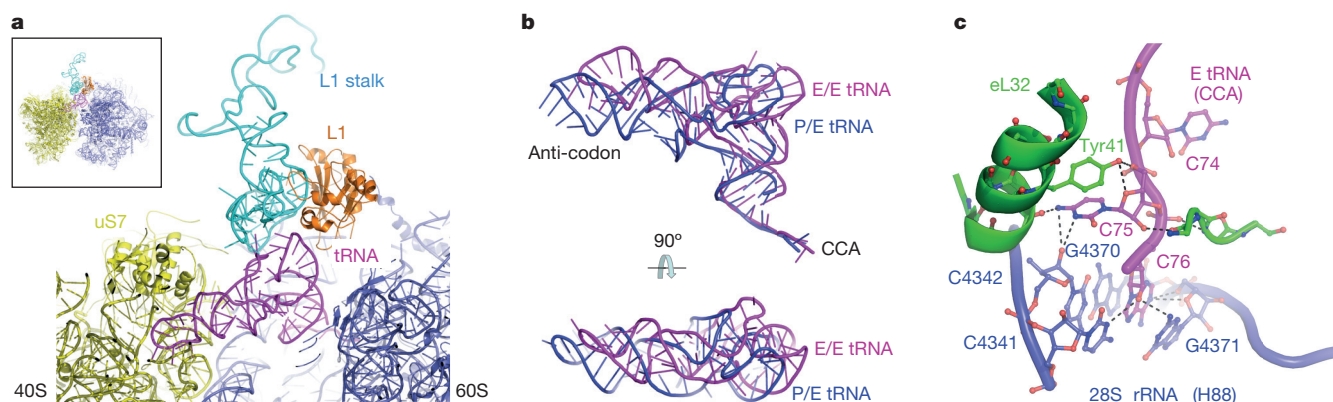

Figure 6 | Molecular recognition by the human ribosome at the level of the E-site tRNA. a, Top view of the E-site (ribosomal subunits and the L1 stalk are annotated). **b**, Comparison of the E-site tRNAs from the human and porcine ribosome structures (refs 1, 8; ribosome in the E/E and P/E states, anticodon and CCA regions are labelled). **c**, Details of the interactions between the CCA end region of the E-site tRNA and the 28S rRNA and protein eL32 of the 60S subunit, showing an intercalation of base C76 in between nucleotides G4370 and G4371, and a π-stacking of base C75 with Tyr41 of eL32. The antibiotic cycloheximide would clash with the CCA region of the tRNA.

protein at the level of the tRNA elbow. The L1 stalk rRNA interacts with the tRNA close to the elbow region and at the top of the decoding stem (Fig. 6a). The CCA end exhibits specific interactions in a ligand pocket, formed by G4370 and G4371 of the 28S rRNA and ribosomal protein eL32 (Fig. 6c). A76 of the tRNA intercalates between bases G4370 and G4371, while C75 forms π-stacking interactions with Tyr41 of eL32 (Fig. 6c). The existence of molecular recognition events in the E-site binding pocket is an interesting observation considering that the tRNA behaves as a reaction product, which maintains specific interactions with the enzyme, after the transfer of the nascent peptide to the new P-site tRNA during the translation process. These specific interactions are present although the tRNA probably is a mixture of co-purified tRNAs, suggesting that these interactions are common to all tRNA, consistent with the universal conservation of the CCA end sequence[40]. The binding pocket of the tRNA CCA end overlaps with the binding site of the antibiotic cycloheximide known from the yeast ribosome[41]. On the 40S side, the E-site tRNA decoding stem fits into a pocket formed by protein uS7 and bases G961, A1641 and C1684 of the 18S rRNA (Extended Data Fig. 6). Although details of the ribosome interactions with the anti-codon cannot be resolved, their presence suggests that interactions with the mRNA are partially released in the E-site as seen previously in 70S crystal structures[42].

While the A- and P-sites are empty in the present structure, the two binding pockets are precisely preformed with many structurally well-ordered side chains for molecular recognition of the tRNA molecules. Exceptions to this are A4548 (28S rRNA helix H93, part of the PTC, corresponds to A2602 in *E. coli*) and the central part of protein uL16 that forms a loop (residues 101–117), which are essentially disordered in the P-site (while the C-terminal part of uL16 located in the A-site is fully defined; Extended Data Fig. 6). This suggests that these components of the P-site pocket actively accommodate the tRNA acceptor stem and CCA end upon binding and become structurally ordered (in the presence of a P-site tRNA the loop interacts with bases 1 and 2 of the tRNA)[43].

The structural stabilization upon substrate binding also applies to helix H69 on which the P-site tRNA decoding stem lodges, and to U4452 in the PTC. H69 shows some structural disorder, while 18S rRNA helix h44 with which it forms bridge B2a is well ordered. Thus, P-site tRNA binding is likely to reinforce inter-subunit interactions across this bridge through the stabilization of H69, consistent with the idea that P-site tRNA binding stabilizes the fully assembled 80S complex, which is required for translation initiation[44–46]. Similarly, nucleotide U4452 of 28S rRNA at the entry of the peptide exit channel is disordered in the absence of growing peptide, while all side chains forming the peptide channel through the 60S subunit are well ordered. U4452 is located at a potential branching point of the peptide channel, suggesting that this residue could serve as a gatekeeper to correctly orient the early growing nascent peptide.

Regarding the A-site, all ribosomal residues including nucleotides U4501, U4500 and G4499 (helix H92 of 28S rRNA) are well defined and pre-oriented for base-pair interactions respectively with the complementary C74, C75 and A76 nucleotides of the tRNA, with U4438 (H89 of 28S rRNA) prepared for recognition of A73 of the tRNA. On the 40S subunit, the decoding centre is entirely pre-formed including most side chains. The universally conserved nucleotide residues A1824 and A1825 of 18S rRNA helix h44 (A1492 and A1493 in *E. coli* numbering) are not resolved well enough to designate their position as flipped-in/out, which is known to be important for decoding[47].

Outlook

This work provides, to our knowledge, the first high-resolution structure of the human 80S ribosome with unprecedented insights into rRNA and amino acid residues, including detailed side-chain positions and conformations. It reveals the basis of molecular recognition through hydrogen bonding between ribosomal RNAs and/or protein moieties within the entire ribosome structure, at the level of the exit site

tRNA and within the subunit interface, which is seen to remodel upon rotation of the ribosomal subunits. Comparison of the pre- and post-translocation states illustrates how inter-subunit bridges have a marked impact on the dynamics of the human ribosome and provides atomic details about subunit interface rearrangements. While the present ribosome complex contains only an E-site tRNA for which detailed inter-actions are described here, its structure constitutes a reference for future functional complexes with ribosomal factors such as mRNA, tRNA and associated protein factors, either directly involved in the translation mechanism or implicated in general translation regulation. Biochemically, the present human 80S ribosome samples will be better suited for forming functional complexes because they do not contain any hibernation or starvation factors such as Stm1 (ref. 2) or serpine 1 mRNA-binding protein 1 (SERBP1)[1], located within the path of the mRNA in previous human and yeast ribosome structures.

Another important aspect is that, for drug design purposes, it is crucial to work with the human target rather than with that of related model systems because of unpredictable effects such as second layer residues around ligand-binding pockets, which can modify the inter-action pattern of the primary contacting residues or open new pockets by conformational changes of side chains. Our structure provides a valuable repertoire and molecular description of antibiotic binding pockets (Fig. 2f) and other new potentially druggable ligand-binding pockets, which can serve as a basis for structure-guided computer-assisted drug design, *in silico* screening and structure determination of complexes with pharmaceutically relevant ligands. In this respect, we foresee two major applications: (1) reducing side effects of current antibiotics by improving drug specificity; and (2) developing new anticancer agents that target the human ribosome to specifically reduce the elevated protein synthesis rate of cancer cells that makes them a primary target over normally growing cells. The present high-resolution human ribosome structure will hopefully provide new opportunities in drug developments directly relevant for human health.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Anger, A. M. *et al.* Structures of the human and *Drosophila* 80S ribosome. *Nature* **497,** 80–85 (2013).
2. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334,** 1524–1529 (2011).
3. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331,** 730–736 (2011).
4. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S. & Ban, N. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334,** 941–948 (2011).
5. Spahn, C. M. *et al.* Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* **118,** 465–475 (2004).
6. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* **13,** 1695–1706 (2005).
7. Chandramouli, P. *et al.* Structure of the mammalian 80S ribosome at 8.7 Å resolution. *Structure* **16,** 535–548 (2008).
8. Voorhees, R. M., Fernandez, I. S., Scheres, S. H. & Hegde, R. S. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* **157,** 1632–1643 (2014).
9. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343,** 1485–1489 (2014).
10. Greber, B. J. *et al.* The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515,** 283–286 (2014).
11. Brown, A. *et al.* Structure of the large ribosomal subunit from human mitochondria. *Science* **346,** 718–722 (2014).
12. Li, X., Zheng, S. Q., Egami, K., Agard, D. A. & Cheng, Y. Influence of electron dose rate on electron counting images recorded with the K2 camera. *J. Struct. Biol.* **184,** 251–260 (2013).
13. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).

14. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2,** e00461 (2013).
15. Khatter, H. *et al.* Purification, characterization and crystallization of the human 80S ribosome. *Nucleic Acids Res.* **42,** e49 (2014).
16. DeLaBarre, B. & Brunger, A. T. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr. D* **62,** 923–932 (2006).
17. Afonine, P. V. *et al.* FEM: feature-enhanced map. *Acta Crystallogr. D* **71,** 646–666 (2015).
18. Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330,** 1203–1209 (2010).
19. Fei, J., Kosuri, P., MacDougall, D. D. & Gonzalez, R. L. Jr. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol. Cell* **30,** 348–359 (2008).
20. Cornish, P. V. *et al.* Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl Acad. Sci. USA* **106,** 2571–2576 (2009).
21. Feng, S., Chen, Y. & Gao, Y. G. Crystal structure of 70S ribosome with both cognate tRNAs in the E and P sites representing an authentic elongation complex. *PLoS ONE* **8,** e58829 (2013).
22. Valle, M. *et al.* Locking and unlocking of ribosomal motions. *Cell* **114,** 123–134 (2003).
23. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406,** 318–322 (2000).
24. Budkevich, T. V. *et al.* Regulation of the mammalian elongation cycle by subunit rolling: a eukaryotic-specific ribosome rearrangement. *Cell* **158,** 121–131 (2014).
25. Budkevich, T. *et al.* Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol. Cell* **44,** 214–224 (2011).
26. Chen, J., Petrov, A., Tsai, A., O'Leary, S. E. & Puglisi, J. D. Coordinated conformational and compositional dynamics drive ribosome translocation. *Nature Struct. Mol. Biol.* **20,** 718–727 (2013).
27. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7,** 499–512 (2001).
28. Klaholz, B. & Moras, D. C.-H. O hydrogen bonds in the nuclear receptor RARgamma–a potential tool for drug selectivity. *Structure* **10,** 1197–1204 (2002).
29. Liu, A. *et al.* NMR detection of bifurcated hydrogen bonds in large proteins. *J. Am. Chem. Soc.* **130,** 2428–2429 (2008).
30. Jiang, L. & Lai, L. C. H. O hydrogen bonds at protein-protein interfaces. *J. Biol. Chem.* **277,** 37732–37740 (2002).
31. Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310,** 827–834 (2005).
32. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N. & Noller, H. F. X-ray crystal structures of 70S ribosome functional complexes. *Science* **285,** 2095–2104 (1999).
33. Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292,** 883–896 (2001).
34. Shenvi, C. L., Dong, K. C., Friedman, E. M., Hanson, J. A. & Cate, J. H. Accessibility of 18S rRNA in human 40S subunits and 80S ribosomes at physiological magnesium ion concentrations–implications for the study of ribosome dynamics. *RNA* **11,** 1898–1908 (2005).
35. Moore, M. N. & Spremulli, L. L. Effects of cations and cosolvents on eukaryotic ribosomal subunit conformation. *Biochemistry* **24,** 191–196 (1985).
36. Sperrazza, J. M., Russell, D. W. & Spremulli, L. L. Reversible dissociation of wheat germ ribosomal subunits: cation-dependent equilibria and thermodynamic parameters. *Biochemistry* **19,** 1053–1058 (1980).
37. Zhang, W., Dunkle, J. A. & Cate, J. H. Structures of the ribosome in intermediate states of ratcheting. *Science* **325,** 1014–1017 (2009).
38. Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332,** 981–984 (2011).
39. Tourigny, D. S., Fernandez, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340,** 1235490 (2013).
40. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313,** 1935–1942 (2006).
41. Garreau de Loubresse, N. *et al.* Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513,** 517–522 (2014).
42. Jenner, L., Rees, B., Yusupov, M. & Yusupova, G. Messenger RNA conformations in the ribosomal E site revealed by X-ray crystallography. *EMBO Rep.* **8,** 846–850 (2007).
43. Rhodin, M. H. & Dinman, J. D. A flexible loop in yeast ribosomal protein L11 coordinates P-site tRNA binding. *Nucleic Acids Res.* **38,** 8377–8389 (2010).
44. Yamamoto, H. *et al.* Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nature Struct. Mol. Biol.* **21,** 721–727 (2014).
45. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83,** 779–812 (2014).
46. Pestova, T. V. Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl Acad. Sci. USA* **98,** 7029–7036 (2001).
47. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484,** 256–259 (2012).

Author Contributions H.K. conducted purification, optimization of samples for cryo-EM and cryo-EM data processing. A.G.M. performed cryo-EM data acquisition, image processing, structure refinement and model building. S.K.N. performed structure refinement and model building. B.P.K supervised the study. All authors analysed the data. B.P.K and H.K. wrote the manuscript, with input from A.G.M. and S.K.N.

METHODS

No statistical methods were used to predetermine sample size.

Purification. Human 80S ribosomes were prepared from HeLa cells, purified to homogeneity and characterized biochemically as described earlier[15]. Briefly, the HeLa cells were lysed and the cytosolic extract was then loaded onto 30% sucrose cushion for overnight centrifugation. The ribosomal pellet was treated with puromycin before the final step of sucrose gradient and PEG precipitation. Concentration of purified 80S was calculated using 1 A260 unit corresponds to 20 pmol of 80S ribosome.

Data collection. Freshly prepared human 80S ribosomes (2.5 μl of a 0.5 mg/ml solution) were applied to 300 mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools, Jena, Germany), blotted with filter paper from both sides in the temperature- and humidity-controlled Vitrobot Mark IV apparatus (FEI, Eindhoven, Netherlands, T = 20 °C, humidity 95%, blot force 5, blot time 0.5 s) and vitrified in liquid ethane pre-cooled by liquid nitrogen. Data were collected on the in-house spherical aberration (Cs) corrected Titan Krios S-FEG instrument (FEI, Eindhoven, Netherlands) operating at 300 kV acceleration voltage and at a nominal underfocus of $\Delta z = -0.6$ to $-4.5\,\mu m$ using the second-generation back-thinned direct electron detector CMOS (Falcon II) 4,096 × 4,096 camera and automated data collection with EPU software (FEI, Eindhoven, Netherlands). The microscope was carefully aligned as well as the Cs corrector. The Falcon II camera was calibrated at nominal magnification of 59,000×. The calibrated magnification on the 14 μm pixel camera was 127,272× resulting in 1.1 Å pixel size at the specimen level. The camera was set up to collect 3 frames (start from the second one) out of 17 possible, plus one total exposure image. Total exposure time was 1 s with a dose of 60 ē Å^{-2} (or 3.5 ē Å^{-2} per frame) (where ē specifies electrons).

Image processing. Before particle picking, stack alignment was performed, which included three frames and total exposure image (total four images in the stack). These four images in the stack were aligned by the whole image motion correction method[13]. Thereafter, an average image of whole stack was used to pick 75,000 particles semi automatically using EMAN2 Boxer[48], and the contrast transfer function of every image was determined using CTFFIND3[49] in the RELION workflow[50]. 2D classification was used to remove bad particles (18,500) and particle sorting was done by 3D classification using 6 classes starting from 56,500 particles, resulting in two dominant classes with 10,000 and 45,000 particles in rotated and non-rotated ribosome conformations, respectively (Extended Data Fig. 7); the subpopulation of the rotated ribosome contains elongation factor eEF2 and the ribosome is in a conformation similar to that of the previous lower-resolution human ribosome structure[1]. For the first steps of refinement, images were coarsened by 2 (C2-images) and 4 (C4-images) using RELION. The initial 2D classification was done with the C4 images, followed by splitting of the different states of the ribosomes using C2 images. Only once the states were clearly separated, uncoarsened (C1) data was used. During the last step of refinement, only close to focus data up to defocus $-1.4\,\mu m$ (about 24,000 particles) was used. At the very last step, the 'movie processing' procedure as implemented in RELION 1.3 version was used with the 3 frames included in the calculations as individual frames. A post-processing procedure implemented in RELION 1.2 (ref. 50) was applied to the final maps for appropriate masking, B-factor sharpening and resolution validation to avoid over-fitting[51]. In this procedure, the appropriate B-factor was determined according to ref. 52. The resolution was estimated in RELION[50], according to the 'gold standard' Fourier shell correlation (FSC = 0.143)[52,53] (Extended Data Fig. 1), indicating an average resolution of 3.6 Å. After calculating local resolution values with ResMap[54] some regions are shown to reach ~2.4 Å resolution in the best cases, and up to 6 Å in the more flexible regions. In general, the 40S subunit appears more flexible resulting in some less well-defined regions at lower resolution as compared to the core of the 60S subunit, which is structurally more stable with the exception of the ES, L1 and P-stalk regions. Map interpretation was done using Chimera[55] and COOT[56] starting from the available atomic models[1,8]. For this, starting atomic models of human and porcine ribosomes[1,8] were placed as individual subunits, rebuilt and

the RNA and protein sequences modified according to the human gene sequence. Most ribosomal proteins have about 95–100% conserved sequences in mammals except eL43, eL29, uL29, eL14 and eL6, which were rebuilt with the human sequences (ribosomal proteins follow the new nomenclature[57]). The atomic model was refined against the experimental map using Phenix[58], which included real space refinement, positional refinement, grouped B-factor refinement and simulated annealing. FEM maps were used for structure interpretation and fine-tuning the positions of side-chain conformations throughout the entire structure by manual model building, followed by simulated annealing protocols. We monitored the refinement process with R-factor values to avoid over-fitting, and refined the entire 80S structure at once (to ensure proper distance restraints between residues throughout the structure, notably at the subunit interface) by simulated annealing using parallel computing, rather than refining each protein subunit individually as for the recent porcine ribosome structure[8]. The R-factor value of the refined structure (35.7%) compares well with that of crystal structures[58] showing that the atomic model is well refined compared to crystallography standards. The final atomic model comprises ~220,000 atoms (excluding hydrogens) across the 5,866 nucleotide residues and ~11,590 amino acids of the 80 proteins and the 4 rRNAs (28S, 5S, 5.8S and 18S; excluding certain expansion segment rRNA which are only partially visible at the periphery of the structure probably due to conformational heterogeneity) and 239 Mg^{2+} ions. The structure extends the completeness of the recent porcine ribosome structure at the level of several regions, such as the position of some rRNA segments (however, most rRNA expansion segments remained poorly defined). This includes the region of the L1 rRNA stalk and the L1 protein that could be completed; albeit at lower local resolution, this region is rather well ordered due to the presence of a tRNA in the E-site, which was fully built in the atomic model. Side chains in various sites and pockets are largely well defined (for example, tRNA sites, PTC, decoding site, peptide channel, mRNA channel), unless they become ordered upon substrate binding (see main text); interestingly, the α-sarcin-ricin loop region (rRNA helix H91) is entirely ordered even in the absence of ribosomal GTPases (including the A4605 region, which is catalytically relevant for GTPase activation and the non base-paired G4600 located next to the G-domain of factors). Protein residues of the final atomic model show well-refined geometrical parameters (allowed regions 22.5%, preferred regions 67.9% and 9.5% of outliers in Ramachandran plots, r.m.s. bond deviations of 0.008Å and angle deviations of 1.24°, which compare well with values observed for refined crystal structures). Figures were prepared using the software Chimera[55] and Pymol[60].

48. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
49. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
50. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
51. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135,** 24–35 (2013).
52. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).
53. van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *J. Struct. Biol.* **151,** 250–262 (2005).
54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11,** 63–65 (2014).
55. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
57. Ban, N. *et al.* A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* **24,** 165–169 (2014).
58. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68,** 352–367 (2012).
59. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313,** 1935–1942 (2006).
60. DeLano, W. L. The PyMOL Molecular Graphic System. (DeLano Scientiic, 2002).
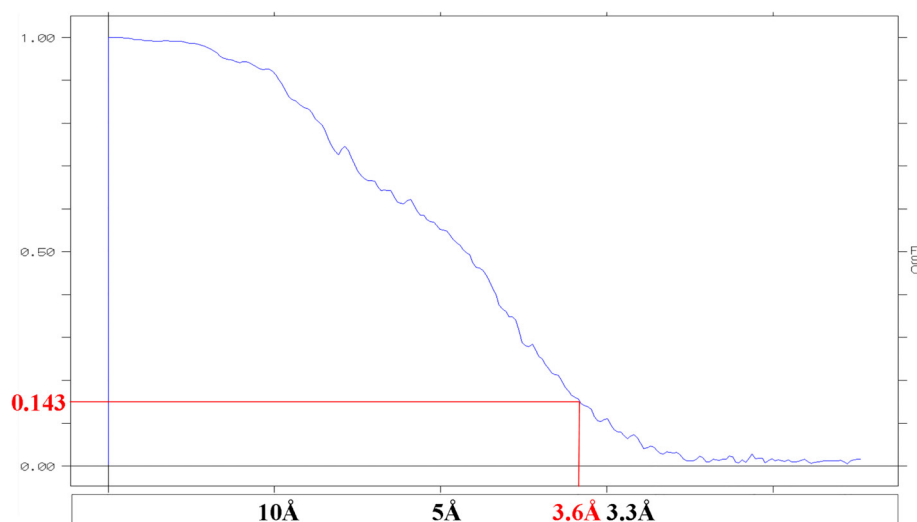
Extended Data Figure 1 | Resolution estimation. The average resolution of the cryo-EM map as estimated from Fourier shell correlation according to the 0.143 criterion.
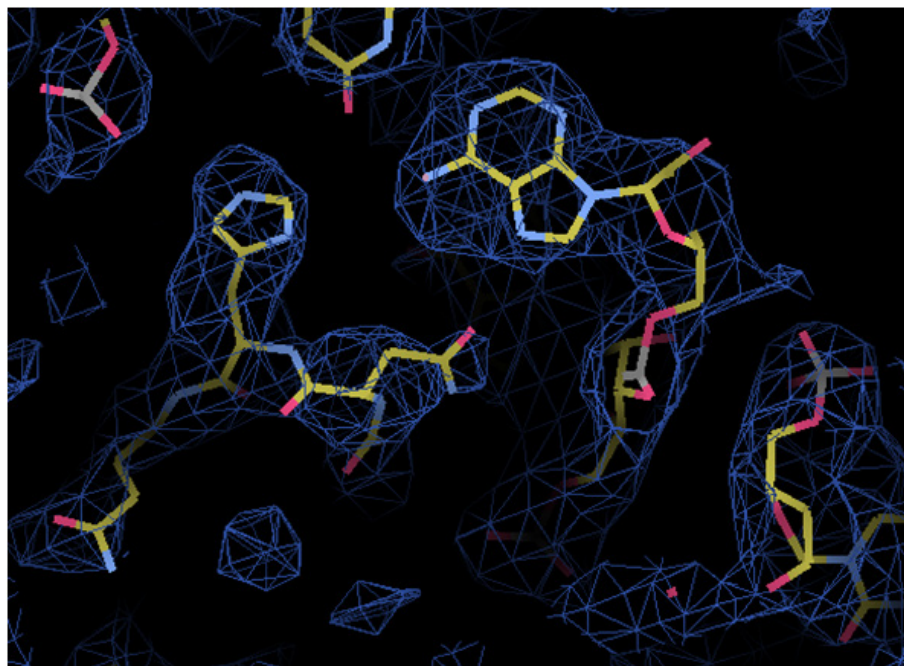
Extended Data Figure 2 | Features of the refined cryo-EM 3D reconstruction. Final cryo-EM map of the peptidyl-transferase centre region illustrating the high level of details observed.

a
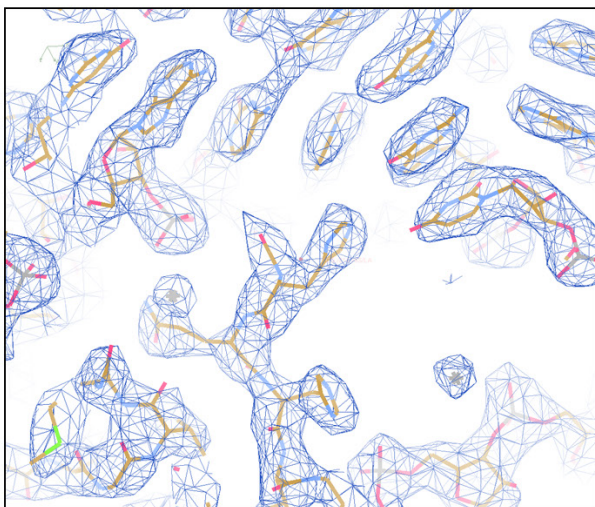
b
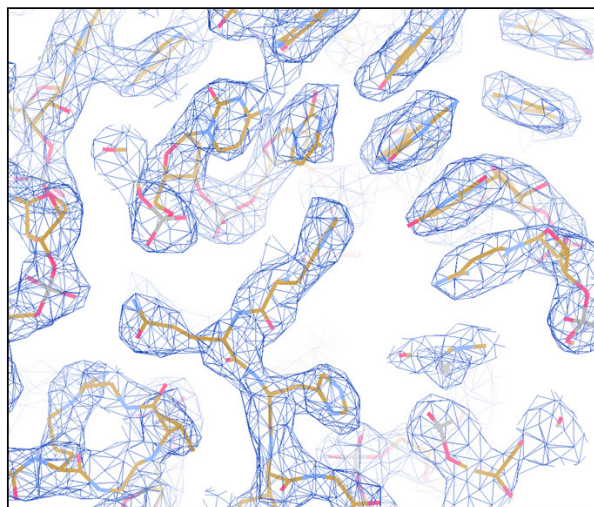
c
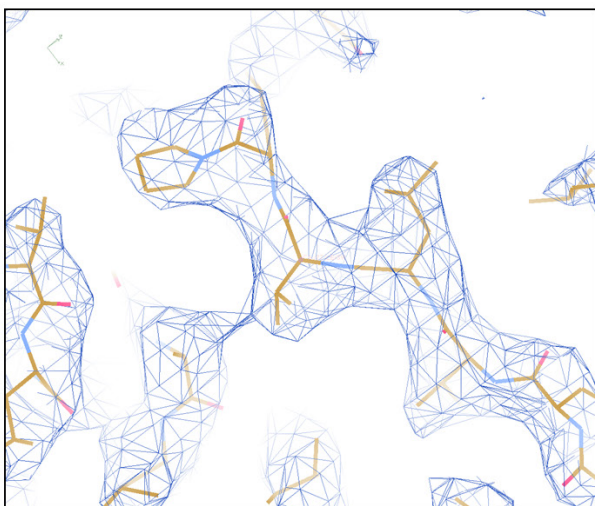
d
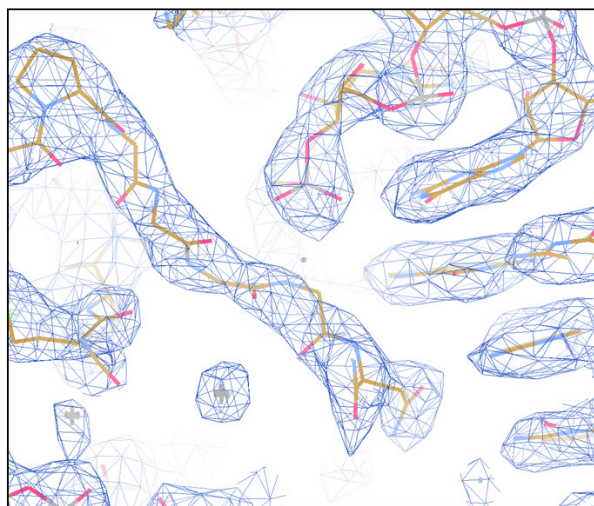
e

f
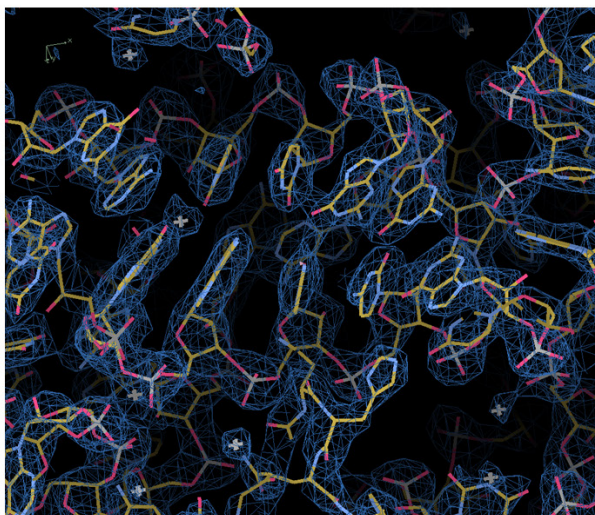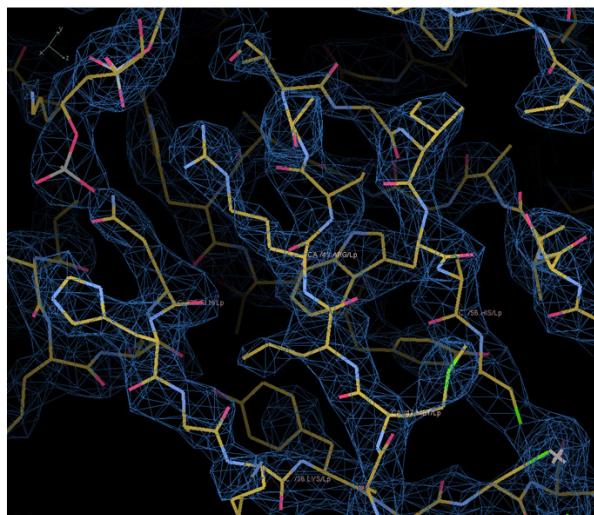
Extended Data Figure 3 | Typical electron density map regions of the human ribosome structure. These final electron density maps were obtained by combining experimental amplitudes derived from the cryo-EM map and phases calculated from the iteratively refined atomic model, as done in standard refinement procedures in X-ray crystallography (see main text).
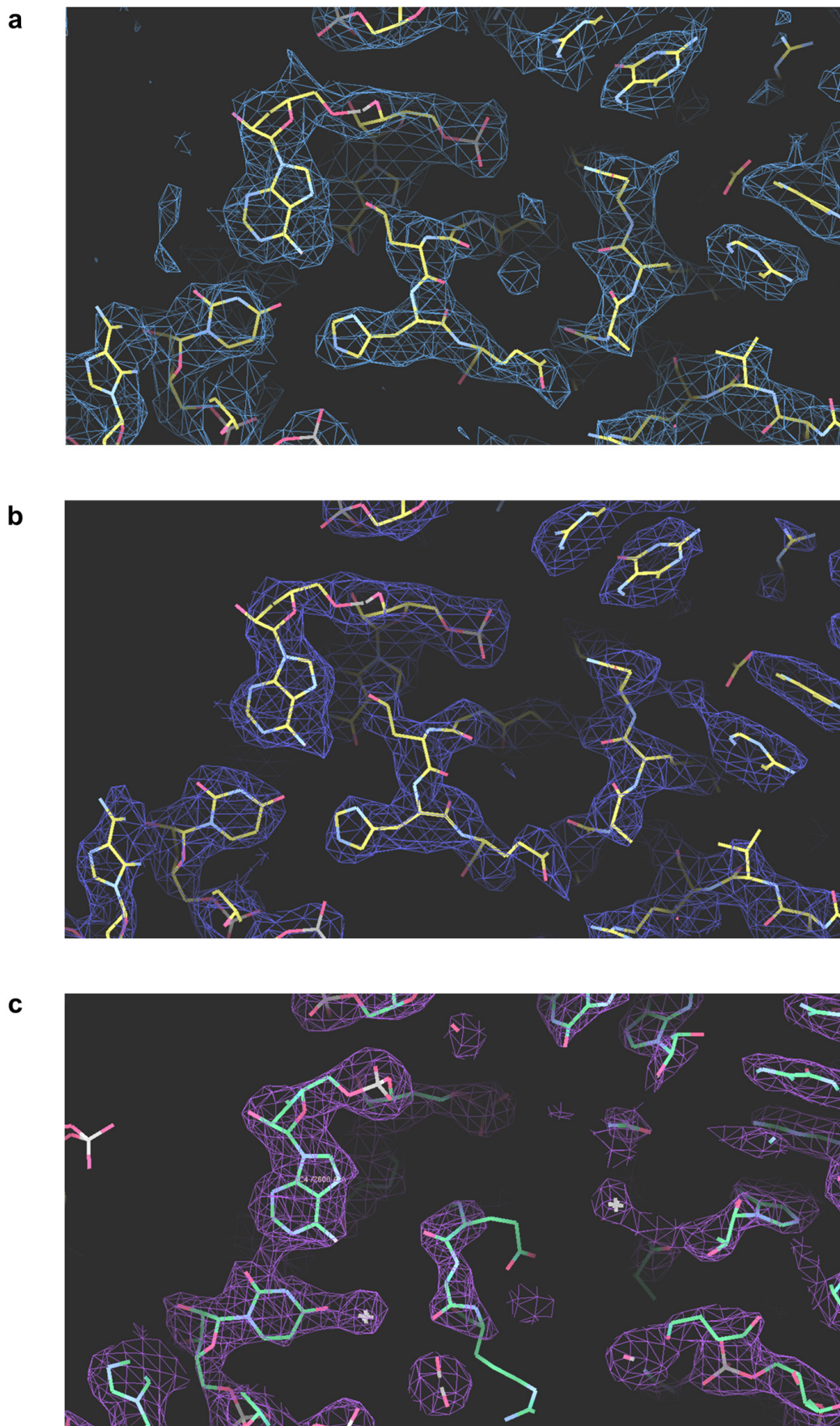
Extended Data Figure 4 | Comparison of maps determined by cryo-EM and X-ray crystallography. Top and middle, cryo-EM map and Phenix map of the human 80S ribosome (this study, A4546 region); bottom, crystal structure of a 70S ribosome at 2.8 Å resolution[59] (corresponding A2600 region).

a

N

C

uS2

b

C

N
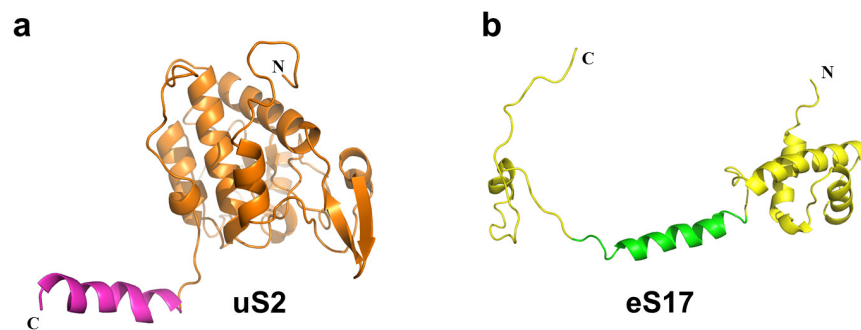
eS17

Extended Data Figure 5 | Ribosomal protein structures. a, b, Structure of ribosomal proteins uS2 and eS17.
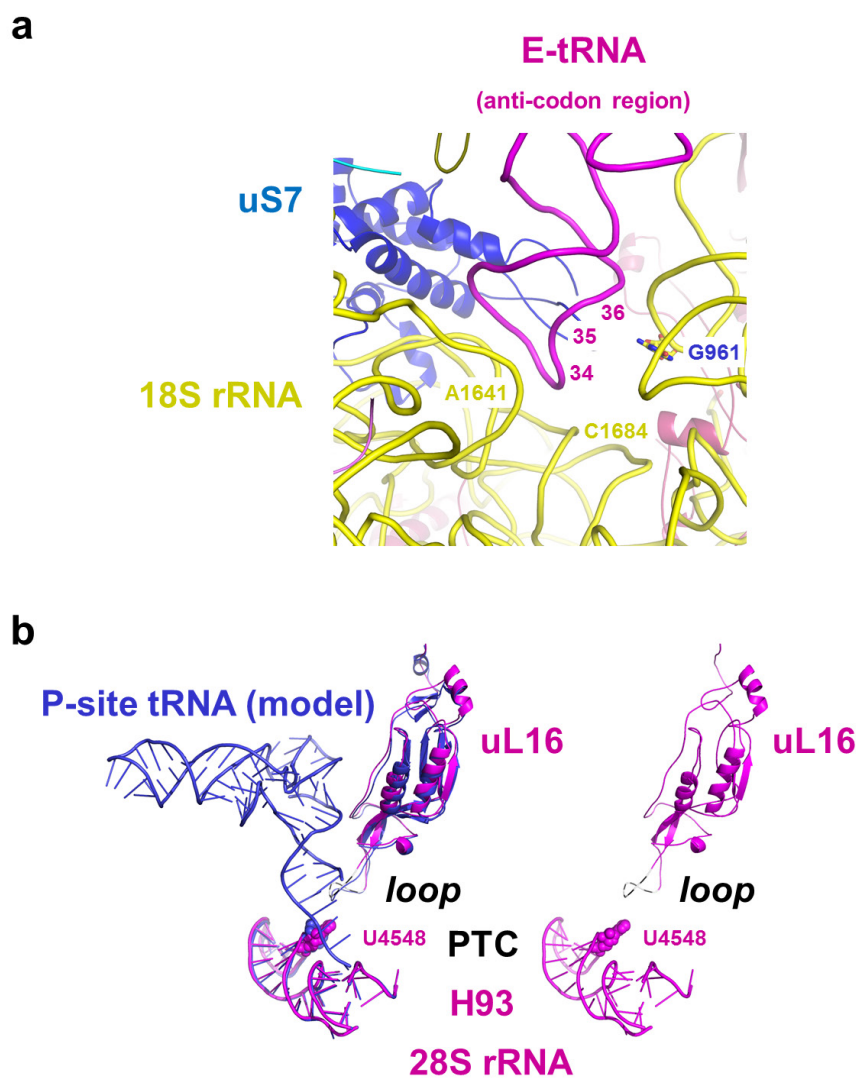
Extended Data Figure 6 | The decoding centre on the 40S. a, Contact regions of the E-site tRNA on the 40S subunit. **b**, Catalytic PTC region of the ribosome, highlighting the partial structural disorder of a region of protein uL16 (loop), and disorder of residue U4548 (28S rRNA) in the absence of P-site tRNA, while all other residues are well-ordered. This suggests that the P-site pocket is largely pre-defined while U4548 and the loop participate in tRNA accommodation.

Initial data set (75 000 particles, pixel size 1.1 Å)

↓

2D classification (18 500 particles were removed)

↓

3D classification into 6 classes

↓

Class 1
Only 60S
1,000 particles

Classes 2 & 3
Low resolution 80S
500 particles

Class 4

10 000 particles
<u>Ratcheted ribosome with eEF2 bound</u>

Classes 5 & 6

45 000 particles
<u>Non-ratcheted ribosome with E-site tRNA</u>

↓

Extraction of low defocus dataset and movie processing (24 000 particles)

Extended Data Figure 7 | Particle sorting scheme. Particle sorting was done by 3D classification using six classes starting from 75,000 particles, resulting in two dominant classes with 10,000 and 45,000 particles, in rotated and non-rotated ribosome conformations respectively; the highest-resolution close-to-focus data set was refined using movie processing (see Methods).

Extended Data Table 1 | Inter-subunit bridges and 60S and 40S residues involved in inter-subunit interactions

	40S subunit		60S subunit	
B1a	uS19	Arg 10	28S	1766
B1b/c	uS13	Gln 10 Arg 14	uL5	Ile 113 Lys 118
B2a	18S	1707-1709 1826-1829 1849-1850 1059 -1058	28S	3765-3768 3759-3760 3762
B2b	18S	1051	28S	3699
B2c	18S	1180	28S	3694-3695
B2e	18S	938-939	eL43	Arg 85
B3	18S	1815-1819 1719-1721	28S	3805-3807 3794-3796 3628-3629 3834
			eL41	Arg 17 Met 20
B4	18S	1028-1029 677-678	28S	1563-1564
B5	18S	1731-1732	28S	2884-2885
B6	18S	1793	eL24	Arg 47
B7a	18S	970	28S	3710-3711 3742
B7b/c	Not observed			
B8	Not observed			
eB8	eS1	Glu 224 Leu 225	eL8	Glu 260 Lys 264
eB11	eS8	Arg 77 Asn 88 Glu 89 Arg 92 Asp105 Gln 167 Gly 169 Lys 170 Arg 200 Arg 205	28S	5026-5027 5024 2893-2894 3610-3611 3598-3599 5016-5017
eB12	18S	871 908-910 914	eL19	Arg 163 Arg 170 Arg 172 Arg 176
eB13	eS6	Ala 33 Thr 34 Arg131 Ile 141 Lys 143 Leu 144 Phe 145 Asn 146 Leu 147 Ser 148 Asp151 Gln 155 Tyr 156	eL24	Lys 69 Lys70 Arg71 Thr 72 Arg 73 Arg 74 Gln 79 Arg 80 Gly 84 Lys 93 Arg 94 Asn 95 Gln 96 Lys 97
	18S	1779-1781 1783 1746 1748-1750 323 327		Arg 101 Arg 105 Glu 106 Lys 117 Lys 124
eB14	18S	1171-1175 1182-1183 1841-1847 1705-1706	eL41	Met 1 Arg 2 Lys 4 Arg 6 Arg 9 Arg 11 Lys 14 Arg 15 Arg 17 Arg 18 Arg 21

The contacting residues within 4 Å **radius** were selected using Contact (CCP4 suite of programs) and verified in the structure.

LETTER

Extended hard–X–ray emission in the inner few parsecs of the Galaxy

Kerstin Perez[1,2], Charles J. Hailey[1], Franz E. Bauer[3,4,5], Roman A. Krivonos[6], Kaya Mori[1], Frederick K. Baganoff[7], Nicolas M. Barrière[6], Steven E. Boggs[6], Finn E. Christensen[8], William W. Craig[6,9], Brian W. Grefenstette[10], Jonathan E. Grindlay[11], Fiona A. Harrison[10], Jaesub Hong[11], Kristin K. Madsen[10], Melania Nynka[1], Daniel Stern[12], John A. Tomsick[6], Daniel R. Wik[13], Shuo Zhang[1], William W. Zhang[13] & Andreas Zoglauer[6]

The Galactic Centre hosts a puzzling stellar population in its inner few parsecs, with a high abundance of surprisingly young, relatively massive stars bound within the deep potential well of the central supermassive black hole, Sagittarius A* (ref. 1). Previous studies suggest that the population of objects emitting soft X-rays (less than 10 kiloelectronvolts) within the surrounding hundreds of parsecs, as well as the population responsible for unresolved X-ray emission extending along the Galactic plane, is dominated by accreting white dwarf systems[2–5]. Observations of diffuse hard-X-ray (more than 10 kiloelectronvolts) emission in the inner 10 parsecs, however, have been hampered by the limited spatial resolution of previous instruments. Here we report the presence of a distinct hard-X-ray component within the central 4 × 8 parsecs, as revealed by subarcminute-resolution images in the 20–40 kiloelectronvolt range. This emission is more sharply peaked towards the Galactic Centre than is the surface brightness of the soft-X-ray population[5]. This could indicate a significantly more massive population of accreting white dwarfs, large populations of low-mass X-ray binaries or millisecond pulsars, or particle outflows interacting with the surrounding radiation field, dense molecular material or magnetic fields. However, all these interpretations pose significant challenges to our understanding of stellar evolution, binary formation, and cosmic-ray production in the Galactic Centre.

The Galactic Centre region is dense with X-ray-emitting objects[6]; it contains the supernova remnant Sagittarius (Sgr) A East, the colliding stellar winds surrounding Sgr A*, the hot plasma of the Sgr A East plume, dozens of magnetic X-ray filaments, and thousands of resolved[7] and unresolved point sources that constitute the Galactic ridge X-ray emisison[3–5,8]. In hard X-rays, the INTEGRAL satellite has detected emission centred within 1′ of the Galactic Centre[9]. However, the spatial resolution of INTEGRAL's IBIS coded aperture mask (12′) has motivated speculation that the emission results not from a single object, but from a collection of the many surrounding X-ray sources[3].

The NuSTAR X-ray observatory[10], which has an effective area extending from 3 to 79 keV and an angular resolution of 18″ (equivalent to 0.7 pc at the Galactic Centre), viewed the Galactic Centre for a total of 281 ks in July, August and October of 2012. The image of the central 12 pc × 12 pc of the Galaxy (Fig. 1) in the 20–40 keV energy band reveals for the first time a faint diffuse emission that is peaked at the Galactic Centre and extends along the Galactic plane. The image is dominated by this feature, whose spectrum and localization within several parsecs of Sgr A* distinguish it from other unresolved X-ray

emission in the Galaxy, and which has no obvious correlation with radio images of the dense molecular gas of the circumnuclear disk[11] or the dust and gas of Sgr A West[12]. The features prominent in soft-X-ray images no longer visibly contribute, with the exception of bright point-like emission from the pulsar wind nebula G359.95−0.04[13] and fainter emission from the X-ray filament G359.97−0.038[14] and the Cannonball[15] neutron star.
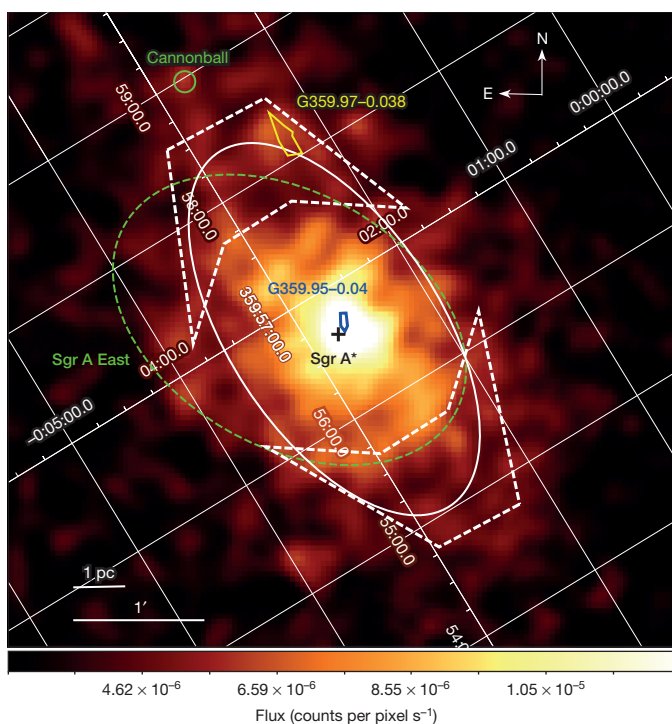

Figure 1 | The 20–40 keV image of the inner 5′ × 5′ (12 pc × 12 pc) of the Galaxy. The colour scale shows flux in units of counts per pixel s^{-1}. The image has been smoothed with a Gaussian kernel of width ~5″ (2 pixels). The solid ellipse (white) illustrates the FWHM of the fit to the unresolved emission. Emission consistent with the pulsar wind nebula G359.95−0.04[13], as well as fainter emission from the Cannonball[19] and non-thermal filament G359.97–0.038[14], is also visible. The dashed ellipse (green) indicates the soft X-ray extent of Sgr A East[19]. Spectra are extracted from the two thick-dashed regions (white polygons).

[1]Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, Room 1027, New York, New York 10027, USA. [2]Haverford College, 370 Lancaster Avenue, KINSC L109, Haverford, Pennsylvania 19041, USA. [3]Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, 306, Santiago 22, Chile. [4]Millennium Institute of Astrophysics, Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile. [5]Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, Colorado 80301, USA. [6]Space Science Laboratory, UC Berkeley, 7 Gauss Way, Berkeley, California 94720, USA. [7]Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 37-555, Cambridge, Massachusetts 02139, USA. [8]DTU Space, National Space Institute, Technical University of Denmark, Elektrovej 327, DK-2800 Lyngby, Denmark. [9]Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94551-0808, USA. [10]Cahill Center for Astronomy and Astrophysics, 1200 East California Boulevard, MC 290-17, California Institute of Technology, Pasadena, California 91125, USA. [11]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-83, Cambridge, Massachusetts 02138, USA. [12]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 4800 Oak Grove Drive, MS 169-221, California 91109, USA. [13]NASA Goddard Space Flight Center, 8800 Greenbelt Road, Code 662, Greenbelt, Maryland 20771, USA.
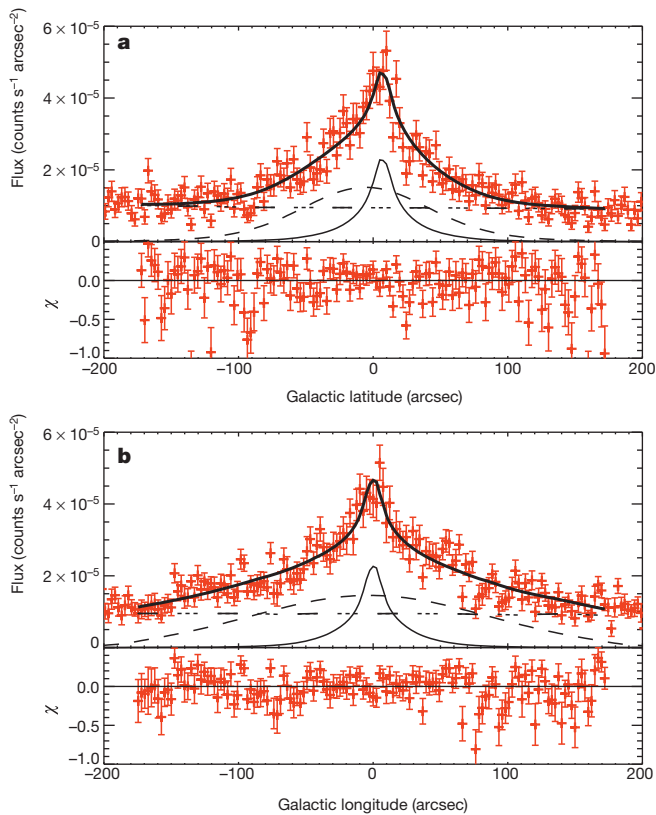
Figure 2 | Profiles of the 20–40 keV data and spatial model along Galactic longitude and latitude. a, Galactic latitude; **b**, Galactic longitude. Top panels, the data (flux in units of counts s^{-1} arcsec^{-2}, red data points; error bars, 1σ) are fitted to a model consisting of an ellipsoidal Gaussian (dashed line), which describes the extended emission, and a point-like Gaussian (thin solid line), which describes G359.95−0.04, both convolved with the on-axis NuSTAR PSF, as well as a background (dash-dot line) that varies between detector chips. Profiles are integrated from 25″ on either side of each axis, with the origin defined at Sagittarius A*. The combined model (thick solid line) and residual emission (lower panel) show that this model describes the data well.
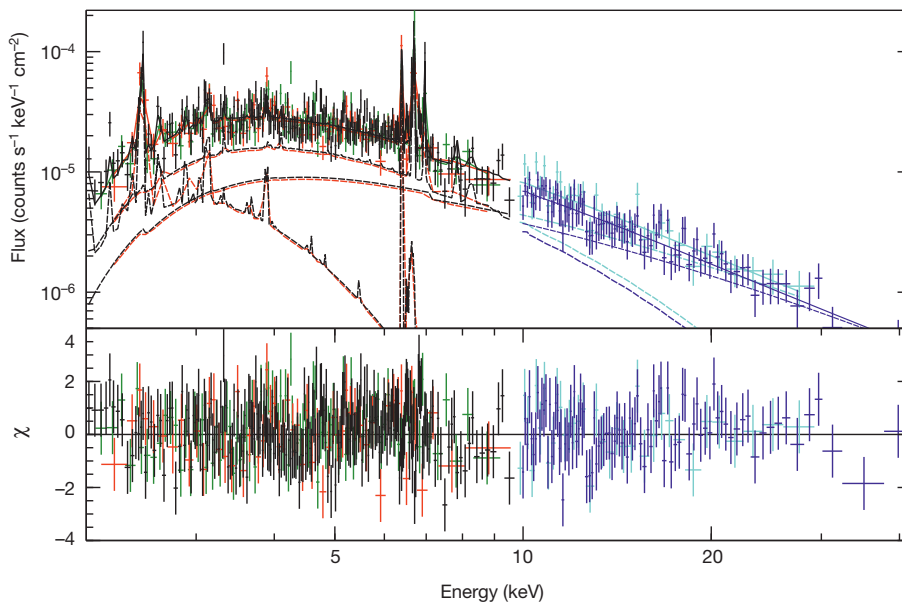
The spatial distribution of this new feature can be described by an ellipsoidal Gaussian with a full-width at half-maximum (FWHM) of $7.8^{+0.9}_{-0.7}$ pc along the Galactic plane and $3.9^{+0.4}_{-0.3}$ pc out of the plane (Fig. 2), where the distance to the Galactic Centre is set to be 8 kpc, the physical extent is quoted before convolution with the NuSTAR point spread function (PSF), and errors are 3σ confidence limits. The centroid of this emission is compatible with the Galactic Centre, defined as the radio location of Sgr A* (right ascension 17 h 45 min 40.036 s, declination −29° 00′ 28.17″; J2000.0)[16], at the 2σ confidence level. As the emission is centred ∼1′ away from the centre of the Sgr A East thermal emission and extends asymmetrically >30″ beyond the super-nova shell to the southwest, a hard-X-ray component of Sgr A East is ruled out as the origin.

We performed spectral analysis in two regions at radii $r \approx 1'$–$2'$ to the southwest and northeast of Sgr A* (see Fig. 1), chosen such that <4% of the emission from G359.95−0.04 contaminates each region (radii are measured from the radio location of Sgr A*). Since the NuSTAR spectral resolution[10] is insufficient to fully resolve the neutral, He-like, and H-like Fe emission lines, we used 2–10 keV XMM-Newton data from these same regions to constrain the temperatures of the plasma emission below 20 keV.

The soft spectrum of the southwest region is dominated by point sources and truly diffuse emission (Fig. 3). Modelling this as two collisional-ionized plasmas[17] yields the best-fit temperatures $kT_1 = 1.0^{+0.3}_{-0.4}$ keV and $kT_2 = 7.5^{+1.6}_{-1.3}$ keV, with a neutral Fe equivalent width of 126^{+46}_{-44} eV at 6.4 keV (90% confidence limits). These values are consistent with those previously obtained by Chandra analysis of faint point sources[2] in the inner 9′ and diffuse emission[18] in the inner $r \approx 2'$–17′, as well as XMM-Newton analysis[5] of the regions $r = 2'$–20′. The $kT \approx 1.0$ keV plasma is attributed to supernova heating of the interstellar medium, coronally active stars, and accreting non-magnetic white dwarfs[4]. The $kT \approx 7.5$ keV plasma is consistent with emission from resolved and unresolved accreting magnetic white dwarfs[2,3,5] with mean white dwarf mass $M_{WD} \approx 0.5 M_\odot$ ($M_\odot$ is the solar mass). The stellar-mass (M) normalized luminosity (L) of our thermal components is $L(2$–10 keV$)/M \approx 9 \times 10^{27}$ erg s^{-1} $M_\odot^{-1}$ at $r \approx 3$ pc, which is consistent with that measured by XMM-Newton[5] at $r \approx 4$ pc. We therefore conclude that our kT_1 and kT_2 components describe the same emission established by previous X-ray observations.

Figure 3 | Unfolded broadband X-ray spectrum of the southwest region. Top panel, the 2–10 keV spectrum consists of XMM-Newton data from the pn (black), MOS1 (red) and MOS2 (green) instruments. Plotted on the y axis is flux in units of counts s^{-1} keV^{-1} cm^{-2}. The 10–40 keV spectrum consists of NuSTAR focal plane A (dark blue) and B (cyan) data. Error bars are 1σ. Above 20 keV, the extended emission dominates, described well by a $kT > 35$ keV bremsstrahlung (here) or a photon index of $\Gamma = 1.3$–1.8 (Extended Data Fig. 4). Dashed lines illustrate separate model components. Bottom panel, fit residuals. Full spectral parameters are given in Extended Data Tables 2 and 3.

Above 20 keV, this thermal emission becomes negligible, and the spectrum is dominated by hard emission from our new feature. This emission can be described by either a power law with photon index $\Gamma = 1.5^{+0.3}_{-0.2}$ and flux $F_{\text{power-law}}(20\text{--}40\text{ keV}) = 6.7 \times 10^{-13}$ erg s^{-1} cm^{-2} or a poorly constrained high-temperature bremsstrahlung at $kT_{\text{high}} = 58^{+127}_{-23}$ keV and flux $F_{kT\text{-high}}(20\text{--}40\text{ keV}) = 6.8 \times 10^{-13}$ erg s^{-1} cm^{-2}. It is critical to note that this hard spectrum cannot be accounted for by extrapolating the previously detected soft thermal emission attributed to accreting magnetic white dwarfs.

The X-ray energy $E_X > 20$ keV spectrum and flux of the northeast region are consistent with those observed in the southwest. The northeast contains significant soft emission from Sgr A East, which dominates over the point source and diffuse populations. Sgr A East is modelled as a two-temperature collisional-ionized plasma, yielding temperatures $kT_1 = 1.1^{+0.1}_{-0.2}$ keV and $kT_2 = 5.1^{+0.9}_{-0.7}$ keV, and metallicity $Z = 2.3^{+0.9}_{-0.4}\,Z_\odot$ (where $Z_\odot$ is the solar metallicity), which are in agreement with Chandra and XMM-Newton measurements[19,20]. Again, this soft thermal emission does not contribute significantly above 20 keV, where the spectrum is fitted well by either a power law with $\Gamma = 1.6^{+0.3}_{-0.4}$ and flux $F_{\text{power-law}}(20\text{--}40\text{ keV}) = 7.6 \times 10^{-13}$ erg s^{-1} cm^{-2} or a high-temperature bremsstrahlung with $kT_{\text{high}} = 66^{+203}_{-30}$ keV and flux $F_{kT\text{-high}}(20\text{--}40\text{ keV}) = 7.4 \times 10^{-13}$ erg s^{-1} cm^{-2}. Faint contaminating emission from the non-thermal X-ray filament G359.97–0.038[14] accounts for the 14% difference in flux between the northeast and southwest regions.

The uniform 20–40 keV spectral and flux values in both regions support the conclusion that the feature is symmetric along the Galactic plane around Sgr A* and is either thermal with $kT_{\text{high}} > 35$ keV or non-thermal with a photon index of $\Gamma = 1.2$–1.9 (90% confidence limits). Using the two-dimensional Gaussian spatial intensity model to scale the flux from the southwest region indicates a luminosity of $L(20\text{--}40\text{ keV}) = 2.4 \times 10^{34}$ erg s^{-1} within the 4 pc $\times$ 8 pc FWHM of emission. This is <10% of the magnitude of the 20–40 keV luminosity of the INTEGRAL Galactic Centre source.

The measured spectrum, luminosity and spatial profile constrain the possible origins of this emission, even given the broad parameter range of our phenomenological model. Truly diffuse sources, such as synchrotron radiation from magnetic filaments[21] or low-surface-brightness pulsar wind nebulae[22], or bremsstrahlung and inverse Compton emission from Sgr A* particle outflows (cosmic rays) interacting with dense molecular material, could produce the observed spectrum. This interpretation is challenging, though, as there is no spatial correlation with corresponding radio images[12] or radiation density models[23], and our current estimates of magnetic filaments and supernova birth rates in this region cannot reproduce the observed luminosity (as described in Methods).

A natural explanation for our emission is provided by the intermediate polar (a type of cataclysmic variable binary star), which has the hardest spectrum of all accreting magnetic white dwarfs. A mean $kT > 35$ keV implies a mean white dwarf mass[24] $M_{\text{WD}} > 0.9 M_\odot$. This is substantially more massive than the population with mean $M_{\text{WD}} \approx 0.5 M_\odot$ previously observed in the Galactic Centre[2,5] and ridge[3], or the $M_{\text{WD}} = 0.66^{+0.09}_{-0.07} M_\odot$ population observed in the Galactic bulge[8]. It is further distinguished from the low-mass intermediate polar population by its surface brightness distribution, which falls more steeply with distance from the Galactic Centre. Refitting our 20–40 keV image in the region $r > 60''$ (to avoid biasing the distribution owing to the presence of G359.95−0.04) shows a decrease with angular offset from Sgr A*, θ, and latitudinal angular offset from Sgr A*, ϕ, of $\theta^{-\alpha}\exp(-|\phi|/\phi_{\text{sc}})$, with $\alpha = 1.4 \pm 0.1$ and latitudinal scale height $\phi_{\text{sc}} = 1.0' \pm 0.1'$ (1σ errors). This is significantly narrower than the $\alpha = 0.6^{+0.02}_{-0.03}$ and $\phi_{\text{sc}} = 18.6'^{+1.6'}_{-1.2'}$ measured for the 2–10 keV distribution[5] at $r > 2'$.

Our observed luminosity implies that about 10^3–10^4 of these massive intermediate polars would exist in the central 4 pc $\times$ 8 pc. Scaling the density of intermediate polars in the solar neighbourhood by the stellar density of the Galactic Centre[25] and the fraction of sufficiently massive progenitor B-stars predicts about three orders of magnitude fewer systems. Though puzzling, this overabundance is of similar magnitude to that implied for the low-mass population[5].

A significant contribution from low-mass X-ray binaries or from millisecond pulsars may also be possible. The stellar density of faint X-ray transients is already known to be higher in the Galactic Centre[26,27]. However, for a population of low-mass black hole[28] or neutron star[29] X-ray binaries to account for the observed luminosity while remaining consistent with SWIFT X-ray outburst monitoring results[26], it must be composed of a rare class of very faint X-ray transients[26,28]. A sufficient population of millisecond pulsars would account for ~40% of the sources identified by Chandra. This in contrast to previous analyses identifying accreting magnetic white dwarfs as the majority[2,4,7], but could indicate a surprising abundance of unidentified millisecond pulsars in the region. This could strengthen the motivation for a millisecond pulsar explanation of the γ-ray excess observed by the Fermi γ-ray space telescope surrounding the Galactic Centre[30].

Regardless of origin, these observations reveal novel processes present within the inner parsecs of the Galaxy. If arising from a population of compact objects, the steeply falling spatial distribution suggests that binary systems are accumulating, forming, and/or experiencing accretion with a much higher probability in the inner parsecs than in the solar neighbourhood or the Galactic bulge. These observations thus provide input useful for studying accretion physics, dynamical formation, and evolution of exotic binaries near the central supermassive black hole. A cosmic-ray origin, in contrast, has implications for models of particle outflow from Sgr A*, as well as radiation and magnetic field distributions.

1. Genzel, R., Eisenhauer, F. & Gillessen, S. The Galactic Center massive black hole and nuclear star cluster. *Rev. Mod. Phys.* **82,** 3121–3195 (2010).
2. Muno, M. P. *et al.* The spectra and variability of X-ray sources in a deep Chandra observation of the Galactic Center. *Astrophys. J.* **613,** 1179–1201 (2004).
3. Krivonos, R. *et al.* Hard X-ray emission from the Galactic ridge. *Astron. Astrophys.* **463,** 957–967 (2007).
4. Revnivtsev, M. *et al.* Discrete sources as the origin of the Galactic X-ray ridge emission. *Nature* **458,** 1142–1144 (2009).
5. Heard, V. & Warwick, R. XMM-Newton observations of the Galactic Centre region — I: The distribution of low-luminosity X-ray sources. *Mon. Not. R. Astron. Soc.* **428,** 3462–3477 (2013).
6. Goldwurm, A. in *The Galactic Center: a Window to the Nuclear Environment of Disk Galaxies* (eds Morris, M. R., Wang, Q. D. & Yuan, F.) Sect. 3 394–401 (Astron. Soc. Pacif. Conf. Ser. Vol. 439, 2011).
7. Muno, M. P. *et al.* A catalog of X-ray point sources from two megaseconds of Chandra observations of the Galactic Center. *Astrophys. J. Suppl. Ser.* **181,** 110–128 (2009).
8. Yuasa, T., Makishima, K. & Nakazawa, K. Broadband spectral analysis of the Galactic ridge X-ray emission. *Astrophys. J.* **753,** 129 (2012).
9. Bélanger, G. *et al.* A persistent high-energy flux from the heart of the Milky Way: INTEGRAL's view of the Galactic Center. *Astrophys. J.* **636,** 275–289 (2006).
10. Harrison, F. A. *et al.* The Nuclear Spectroscopic Telescope Array (NuSTAR) high-energy X-ray mission. *Astrophys. J.* **770,** 103 (2013).
11. Christopher, M. H., Scoville, N. Z., Stolovy, S. R. & Yun, M. S. HCN and HCO$^+$ observations of the Galactic circumnuclear disk. *Astrophys. J.* **622,** 346–365 (2005).
12. Zhao, J.-H., Morris, M. R. & Goss, W. M. in *The Galactic Center: Feeding and Feedback in a Normal Galactic Nucleus* (eds Sjouwerman, L. O., Lang, C. C. & Ott, J.) 364–368 (IAU Symp. Vol. 303, 2014).
13. Wang, Q. D., Lu, F. J. & Gotthelf, E. V. G359.95−0.04: an energetic pulsar candidate near Sgr A*. *Mon. Not. R. Astron. Soc.* **367,** 937–944 (2006).
14. Nynka, M. *et al.* G359.97–0.038: a hard X-ray filament associated with a supernova shell-molecular cloud interaction. *Astrophys. J.* **800,** 119 (2015).
15. Nynka, M. *et al.* High-energy X-rays from J174545.5−285829, the Cannonball: a candidate pulsar wind nebula associated with Sgr A East. *Astrophys. J.* **778,** L31 (2013).
16. Menten, K. M., Reid, M. J., Eckart, A. & Genzel, R. The position of Sagittarius A*: accurate alignment of the radio and infrared reference frames at the Galactic Center. *Astrophys. J.* **475,** L111–L114 (1997).
17. Smith, R., Brickhouse, N., Liedahl, D. & Raymond, J. Collisional plasma models with APEC/APED: emission-line diagnostics of hydrogen-like and helium-like ions. *Astrophys. J.* **556,** L91–L95 (2001).

18. Muno, M. *et al.* Diffuse X-ray emission in a deep Chandra image of the Galactic Center. *Astrophys. J.* **613,** 326–342 (2004).
19. Park, S. *et al.* A candidate neutron star associated with Galactic Center supernova remnant Sagittarius A East. *Astrophys. J.* **631,** 964–975 (2005).
20. Sakano, M., Warwick, R. S., Decourchelle, A. & Predehl, P. XMM-Newton observations of Sagittarius A East. *Mon. Not. R. Astron. Soc.* **350,** 129–139 (2004).
21. Zhang, S. *et al.* High-energy X-ray detection of G359.89-0.08 (Sgr A-E): magnetic flux tube emission powered by cosmic rays? *Astrophys. J.* **784,** 6 (2014).
22. Johnson, S. P., Dong, H. & Wang, Q. D. A large scale survey of X-ray filaments in the Galactic Center. *Mon. Not. R. Astron. Soc.* **399,** 1429–1440 (2009).
23. Davidson, J. *et al.* The luminosity of the Galactic Center. *Astrophys. J.* **387,** 189–211 (1992).
24. Suleimanov, V., Revnivtsev, M. & Ritter, H. RXTE broadband X-ray spectra of intermediate polars and white dwarf mass estimates. *Astron. Astrophys.* **435,** 191–199 (2005).
25. Merrit, D. (ed.) *Dynamics and Evolution of Galactic Nuclei* Ch. 5 276 (Princeton Series in Astrophysics, Princeton Univ. Press, 2013).
26. Degenaar, N. *et al.* A four-year XMM-Newton/Chandra monitoring campaign of the Galactic centre: analysing the X-ray transients. *Astron. Astrophys.* **545,** A49–A73 (2012).
27. Muno, M. P. *et al.* An overabundance of transient X-ray binaries within 1 pc of the Galactic Center. *Astrophys. J.* **622,** L113–L116 (2005).
28. Menou, K., Narayan, R. & Lasota, J.-P. A population of faint nontransient low-mass black hole binaries. *Astrophys. J.* **513,** 811–826 (1999).
29. Heinke, C. *et al.* Discovery of a second transient low-mass X-ray binary in the globular cluster NGC 6440. *Astrophys. J.* **714,** 894–903 (2010).
30. Hooper, D. & Linden, T. On the origin of the gamma rays from the Galactic Center. *Phys. Rev. D* **84,** 123005 (2011).

Author Contributions K.P.: image analysis, spectral analysis, modelling of the Sagittarius A* region, interpretation and manuscript preparation. C.J.H.: interpretation, manuscript preparation and review. F.E.B.: interpretation and manuscript review. R.K.: image analysis and interpretation. K.M.: interpretation, manuscript preparation and review. F.K.B., N.M.B., S.E.B., J.E.G., J.H. and J.A.T.: interpretation and manuscript review. F.E.C., W.W.Z.: optics production and calibration. W.W.C.: optics and instrumentation production and response, and observation planning. B.W.G., K.K.M., D.R.W. and A.Z.: software for data analysis, background modelling, and calibration. F.A.H.: NuSTAR principal investigator, observation planning, interpretation and manuscript review. M.N., S.Z.: data preparation and interpretation. D.S.: observation planning and manuscript review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.P. (kperez1@haverford.edu).

METHODS

We use data from three NuSTAR observations of the Sgr A* region in 2012: 20–23 July (154 ks), 4–6 August (77 ks), and 16–18 October (50 ks). These correspond to the observation IDs 30001002001, 30001002003, and 30001002004, respectively. Data reduction and spectral extraction were performed with the NuSTAR Data Analysis Software (NuSTARDAS) v1.1.0 software package. The HEASoft v6.13 package was used for data analysis.

Event selection. We removed all data from within 100 s of passage through the South Atlantic Anomaly and time periods corresponding to Sgr A* flares, as observed by NuSTAR or during coincident Chandra observations[31]. Data from focal plane B of the July and October observations are not included due to stray light[10] from the nearby bright objects GX 3+1 and 1E1740-2942. After data screening and combining available focal plane A (FPMA) and B (FPMB) observations, the analysis contains ~297 ks of effective exposure time.

Image analysis. The nominal NuSTAR image coordinates are accurate to 8″ (90% confidence level)[10]. To improve our positional accuracy, we derived an astrometric correction for each observation by registering known sources. For the July and October observations, we used the radio position of Sgr A*[16], the NuSTAR position of the Sgr A* flare in the 3–79 keV band, the Chandra position of Sgr A-E[22], and the NuSTAR position of Sgr A-E[21] in the 3–10 keV band for source registration. For the August observation, where the lack of a flare makes localization of Sgr A* difficult with NuSTAR, we instead used the Chandra position of the Cannonball[19], the NuSTAR position of the Cannonball[15] in the 10–40 keV band, the Chandra position of Sgr A-E, and the NuSTAR position of Sgr A-E in the 3–10 keV band.

After each photon event file was position corrected, we combined NuSTAR images in sky coordinates and normalized the resulting image by an effective exposure map that accounts for observation time, but not vignetting effects. Since the optical axis was near Sgr A* for all observations and detector background rates in the 20–40 keV energy band are non-negligible, vignetting was accounted for in the spatial fitting by convolving the source components, but not detector background components, with the on-axis NuSTAR point spread function (PSF). We defined a systematic error on the astrometric correction to be the maximum distance between the reference positions and the NuSTAR centroid positions of the Cannonball and Sgr A-E knot in the resulting image, yielding uncertainties of 1.0″ in RA and 1.2″ in Dec. This is in addition to the 4″ spatial uncertainty from the PSF smearing effect due to incomplete aspect reconstruction[32].

We used the Sherpa[33] package to fit a phenomenological model to the 20–40 keV raw photon count sky image. Due to low source counts, the C-statistic was used. This model was composed of one symmetric point-like Gaussian and one ellipsoidal extended Gaussian source, both convolved with the on-axis NuSTAR PSF stored in the NuSTARDAS v1.1.0 CALDB database, and a flat internal background that varies between detector chips. The background map was derived using blank sky observations to fix the relative normalization of the detector background between each chip. In this energy range, the stray light background due to the cosmic X-ray background and Galactic ridge X-ray emission are subdominant compared to the internal background. The overall normalization of the background map was left as an independent fit parameter. The total effective exposure map was used to normalize all three fit components. The final model used for fitting was thus:

$$\text{Psf} \times (\text{gauss2d.sourceOne} + \text{gauss2d.sourceTwo}) * \text{emap}$$
$$+ \text{bkgdmap} * \text{emap}$$

where psf is the on-axis NuSTAR PSF, gauss2d.sourceOne is the point-like Gaussian, gauss2d.sourceTwo is the ellipsoidal extended Gaussian, bkgdmap is the background map, and emap is the effective exposure map. Here, * indicates multiplication and × indicates convolution. To minimize variations in background and detector response across different detector chips, and to avoid emission from the molecular cloud regions to the northeast, the fit was performed within a radius $r < 3′$ from the Galactic Centre.

The 20–40 keV model and residuals in the fit region are shown in Extended Data Fig. 1. No remaining significant structure is present in the residual map. The component gauss2d.sourceOne fits to a FWHM of $1.8^{+2.3}_{-0.7}{}''$, where 2.5″ is one image pixel. This is a lower limit on the spatial extent, consistent with a source small compared to the 18″ (FWHM) NuSTAR PSF. The second component fits to a Gaussian with FWHM = $195.8^{+21.9}_{-16.8}{}''$ along the major axis, FWHM = $101.8^{+11.4}_{-8.7}{}''$ along the minor axis, and an inclination of $\theta = 57°$ from the positive northern axis. The centroid of this component is located at $\Delta\text{RA} = -2.2^{+4.5}_{-3.8}{}''$ and $\Delta\text{Dec.} = 2.8^{+3.4}_{-4.1}{}''$ from Sgr A*. Because calculating fit errors while leaving all parameters free was too computationally intensive, the amplitude of the background map and the extended component ellipticity and θ were fixed during error calculation. All fit parameters are listed in Extended Data Table 1. Consistent

values of centroid location and spatial extent are obtained by fitting a single ellipsoidal Gaussian and background model in a region that excludes the central 60″ radius, thus minimizing emission from G359.95–0.04.

To compare the NuSTAR 20–40 keV morphology with the 2–10 keV surface brightness distribution, we re-fitted the NuSTAR data with the spatial model used by the Heard and Warwick[5] analysis of XMM-Newton data in the region $r = 2′–20′$:

$$S = N\theta^{-\alpha}\exp(-|\phi|/\phi_{sc}) + \text{const.} \qquad (1)$$

This model describes a surface brightness S falling as a function of the angular offset θ (measured in arcminutes) from the position of Sgr A* and the latitude angle ϕ, measured with respect to the latitude of Sgr A*. The parameters N, α and ϕ_{sc} represent the normalization of the power law, the power-law index, and the latitudinal scale height in arcminutes, respectively.

To avoid biasing the steepness of the distribution, we excluded the central $r < 60″$ region that contains the bright pulsar wind nebula G359.95−0.04. We also exclude the emission from $r < 20″$ surrounding the Cannonball[15]. The constant background term const. is obtained by fitting the same background map as used in our two-Gaussian phenomenological model to the region $r = 300″–400″$, yielding the same background normalization as obtained in the two-Gaussian model above. We note that this const. term does not represent the true underlying sky surface brightness, as it contains a large contribution from instrumental background. Fitting in the region $r = 60″–180″$ constrains the spatial parameters to be $\alpha = 1.4 \pm 0.1$ and $\phi_{sc} = 1.0′ \pm 0.1′$ (1σ errors). This is significantly narrower in both longitude and latitude than the soft distribution observed by Heard and Warwick, who obtained $\alpha = 0.6^{+0.02}_{-0.03}$ and $\phi_{sc} = 18.6^{+1.6}_{-1.2}{}'$. The effect of vignetting is not important this close to the optical axis. Re-fitting the data using an exposure map that includes the effect of vignetting, calculated at 25 keV, yields consistent values of $\alpha = 1.2 \pm 0.1$ and $\phi_{sc} = 1.0′ \pm 0.1′$.

In order to fit over the full $r < 180″$ region and to further compare with our two-Gaussian model, we fix the magnitude and position of the point-like Gaussian describing G359.95−0.04 to the same values as listed in Extended Data Table 1. Re-fitting the distribution S in this full region yields $\alpha = 1.0′ \pm 0.1′$ and $\phi_{sc} = 1.1′ \pm 0.1′$. The profiles of the total 20–40 keV emission along Galactic longitude and latitude, along with this best-fit model result, are shown in Extended Data Fig. 2.

Spectral analysis. Using the event files with astrometric corrections applied, we extracted source spectra from the dashed regions shown in Fig. 1. Since the 20–40 keV NuSTAR image is dominated by internal detector background, the background spectra were extracted from regions located on the same detector chip that contains the Sgr A* region, but outside the molecular cloud regions (see Extended Data Fig. 3).

We used 2–10 keV data from the XMM-Newton observations carried out on 2012 August 31 (observation ID 0694640301) and 2012 September 24 (observation ID 0694641101), reprocessed with SAS v12.0.1. No bright X-ray outbursts occurred during these periods. We extracted EPIC-pn and EPIC-MOS spectra from the regions used for the NuSTAR analysis and followed the standard data screening procedures from the September 2013 XMM-Newton ABC guide.

The resulting data from the MOS1, MOS2, and pn instruments sum to ~210 ks of effective exposure time. For MOS1, the 2012 September 24 observation fell on a dead detector chip and is not used. Since the XMM-Newton spectra are dominated by soft diffuse emission that fills the central arcminutes, the background spectra were extracted from dimmer regions located at least 10′ to the northwest of the Sagittarius A* region.

Spectral fitting and flux derivations were performed in XSPEC version 12.8.0[34], with photoionization cross-sections as defined in ref. 35 and abundances for the interstellar absorption as defined in ref. 36. All spectra were re-binned so that each spectral bin contained at least 20 counts above the estimated background level, obtained by scaling the background spectrum to the source extraction area.

To analyse the southwest and northeast region spectra of XMM-Newton and NuSTAR, we compared the models const.*tbabs*(apec + apec + gaus + po) and const.*tbabs*(apec + apec + gaus + bremss), where const. is a constant factor, tbabs is the neutral hydrogen absorption model, apec is a collisionally ionized plasma model, gaus is a Gaussian emission line model centred at 6.4 keV with width fixed to $\sigma = 0.01$ keV (to model neutral Fe emission), po is a power-law model, and bremss is a thermal bremsstrahlung model. The final spectral fit values are not significantly sensitive to the width of the Gaussian fixed at 6.4 keV. The thermal bremsstrahlung model was chosen to describe possible very-high-temperature (>30 keV) emission, where emission lines would no longer be present. To account for different normalizations between instruments, the constant factor was allowed to fit freely to each data set.

The best-fit parameters and 90% confidence level statistical errors are listed in Extended Data Tables 2 and 3. In the southwest region, the lower-temperature component has a best-fit metallicity that is very high, $Z = 5Z_\odot$, but varying this parameter has no effect on the spectral parameters of the higher-energy compo-

nents. The higher-temperature component fits to a metallicity $Z = 1.7Z_\odot$. This is consistent with emission from unresolved magnetic accreting white dwarfs, with slightly enhanced metallicity because the lighter elemental abundances, which are measured to be $Z > Z_\odot$ in the Galactic Centre[18], are linked to the Fe abundance in the apec model. Because a degeneracy exists between the values for the abundance of the kT_2 component, the normalization of the kT_2 component, and the normalization of the power-law component, the kT_2 abundance was fixed to its best-fit value during the calculation of parameter error intervals. The hard component can be modelled equally well by either a thermal bremsstrahlung, as shown in Fig. 3, or a power law, as shown in Extended Data Fig. 4. In the northeast region, the hard emission can be described by these same models, as shown in Extended Data Figs 5 and 6. The non-thermal spectrum of G359.97−0.038[14] could contribute at most 19% of the flux observed in the northeast region (at the 90% confidence level).

Spectral analysis of central $r = 20''$. Our spatial model and the asymmetric surface brightness distribution indicate that our observation cannot be attributed to PSF leakage from a bright, central source. To further support this claim, we perform spectral analysis of the $r = 20''$ region surrounding the point-like source attributed to G359.95−0.04. The radius was chosen to maximize the ratio of the point-source flux to the underlying extended flux. The software versions, background regions, and spectral re-binning are identical to those used above.

We first fit the NuSTAR data in the energy range 20–60 keV with a power-law model, po, yielding $\Gamma = 2.2^{+0.4}_{-0.3}$. This is significantly softer than the emission measured in the southwest or northeast regions. To better constrain the photon index, we again combine the 2–10 keV XMM-Newton data with 10–60 keV NuSTAR data. The soft emission is dominated by Sgr A East, with best-fit temperatures and abundances consistent with Chandra and XMM-Newton measurements of this region[19,20]. The hard emission is described by $\Gamma = 2.0^{+0.2}_{-0.2}$, again significantly softer than the best-fit values of the extended emission.

Population analysis. *Truly diffuse sources.* Chandra has identified 17 X-ray-emitting non-thermal filaments in the central $20' \times 20'$ region[22], with photon indices ranging within $\Gamma \approx 1.0$–2.0. Extrapolating the combined 2–10 keV luminosity of the six filaments (excluding G359.95−0.04) that lie within the central 4 pc × 8 pc region using our best-fit $\Gamma = 1.6$, we see that they could contribute at most 30% of the observed luminosity. Also, no X-ray filaments are identified in our southwest region. The faintest filament observed by Chandra has a luminosity of $L(2\text{–}10\,\text{keV}) = 2.0 \times 10^{32}\,\text{erg s}^{-1}$. Interpreting this as a minimum Chandra detection threshold and again extrapolating using the best-fit photon index implies that ~100 additional filaments would be necessary to account for the total luminosity. Radio emission, which is observed from ~25% of X-ray filaments[37], would thus be expected to trace our observed morphology. However, no such similar structure is observed in recent 6 cm radio maps[12].

A population of low-surface-brightness pulsar wind nebulae[38] could also contribute. Again assuming $L(2\text{–}10\,\text{keV}) = 2.0 \times 10^{32}\,\text{erg s}^{-1}$ is the minimum luminosity at which Chandra would detect an extended object, ~100 such objects are required. However, simulations indicate that only a few dozen should exist, spread over the central $r = 20'$ of the Galaxy[22]. Thus an order of magnitude higher supernova birth rate in the Galactic Centre than currently assumed would be required to explain our observations.

In addition to particle outflows from Sgr A*, cosmic rays from shocked stellar winds combined with the high radiation density of the region could produce an inverse Compton origin. To account for the observed 20–40 keV luminosity of the southwest region, which is at a distance of $r \approx 2$–5 pc, a cosmic-ray density similar to that of the central $r < 1$ pc and a radiation density of ~10,000 eV cm^{-3} would be required[39]. However, the radiation density decreases rapidly with distance from Sgr A*[23], falling from ~5,000 eV cm^{-3} in the inner parsec to ~ 50 eV cm^{-3} at a distance of 10 pc, and the cosmic-ray density is also expected to decrease rapidly away from the Galactic Centre.

A more exotic origin of electrons could be the annihilation of dark-matter particles. However, the electron energies required to produce the observed spectrum are too large to be produced by even the heaviest dark-matter candidates[40]. The NuSTAR emission is also much more centrally localized and asymmetric than would be predicted for dark-matter annihilation in the Galactic halo[41].

Intermediate polars. Our estimate of the necessary number of intermediate polars to reproduce the observed spectrum is based on assumed[8] 2–10 keV luminosities ranging from $L_{min} \approx 10^{30}$–$10^{31}\,\text{erg s}^{-1}$ to $L_{max} \approx 10^{33}\,\text{erg s}^{-1}$, with an integral source luminosity distribution $N(>L) \propto L^{-\alpha}$ and spectral index $\alpha \approx 1$–1.5.

The bremsstrahlung model we have used in our spectral fitting approximates the spatially dependent temperature of the cooling accretion column of an intermediate polar by an average colour temperature. To directly determine the mean mass limit implied by our observations, we refit the spectra of the southwest and

northeast regions using the model of ref. 24. Since our soft components agree with previously measured values, we constrain these parameters to stay within their 90% confidence intervals. This yields a 90% confidence level lower limit on the mean white dwarf mass of $M_{WD} > 0.9M_\odot$. As with the high-temperature bremsstrahlung model, the upper limit is not well constrained.

Low-mass X-ray binaries. Our knowledge of the luminosity of quiescent black hole low-mass X-ray binaries is limited, based on a sample of only 17 systems[42]. For typical luminosities of $L(2\text{–}10\,\text{keV}) \approx (2\text{–}4) \times 10^{31}\,\text{erg s}^{-1}$, ~$10^3$ such systems would be required inside the central 4 pc × 8 pc to account for our observed luminosity. In the last decade, X-ray monitoring surveys[26,43] are believed to have uncovered virtually all transient systems within the inner 50 pc with recurrence times of less than ~5–10 yr, outburst durations longer than a few days, and outburst luminosities $L(2\text{–}10\,\text{keV}) > \sim 10^{34}\,\text{erg s}^{-1}$. Assuming a black hole binary outburst recurrence time of $T_r \approx 50$–100 yr and durations ranging from weeks to months[28] and using the integrated Swift monitoring time implies ~12–50 transient events should have been detected from an underlying population of ~10^3 black hole systems. However, only two new transient sources that could plausibly be attributed to low-mass black hole systems have been observed in that time. A large population of neutron star low-mass X-ray binaries is also inconsistent with X-ray monitoring surveys, assuming a mean $L(2\text{–}10\,\text{keV}) \approx 10^{32}\,\text{erg s}^{-1}$ and recurrence times[29] of ~5–10 yr.

Millisecond pulsars. Some millisecond pulsars exhibit non-thermal emission with a typical photon index $\Gamma \approx 1$–2 (ref. 44), consistent with our observed spectrum. To study the X-ray luminosities of these objects, we selected 125 millisecond pulsars from the Australia Telescope National Facility Pulsar Catalog[45] with spin periods and dipole magnetic field strengths consistent with the observed hard-X-ray emitting millisecond pulsars, $P < 10$ ms and $B < 10^{11}$ G. The spin-down power of the selected objects, $\dot{E}$, ranges from 4×10^{32} to $2 \times 10^{36}\,\text{erg s}^{-1}$, with a mean $\dot{E} = 6 \times 10^{34}\,\text{erg s}^{-1}$. Using the relation between non-thermal luminosity and spin-down power established by previous studies[46], $L(2\text{–}10\,\text{keV}) \approx 10^{-4}\dot{E}$, we obtain a mean non-thermal X-ray luminosity of $L(2\text{–}10\,\text{keV}) = 6 \times 10^{30}\,\text{erg s}^{-1}$. Approximately 3,000 millisecond pulsars with this mean X-ray luminosity would be required to account for the our observations. Over 100 of these sources would have a luminosity of $L(2\text{–}10\,\text{keV}) > \sim 4 \times 10^{31}\,\text{erg s}^{-1}$, which is the Chandra point-source detection limit for this region[7]. This implies that over 40% of the sources identified by Chandra within the central 4 pc × 8 pc are millisecond pulsars, a claim that requires further study and is inconsistent with the intermediate polar interpretation of the majority of these sources[2,4,7].

Sample size. No statistical methods were used to predetermine sample size.

31. Barrière, N. M. *et al.* NuSTAR detection of high-energy X-ray emission and rapid variability from Sagittarius A* flares. *Astrophys. J.* **786,** 46 (2014).
32. An, H. *et al.* In-flight PSF calibration of the NuSTAR hard X-ray optics. *Proc. SPIE* **9144,** 1–10 (2014).
33. Fruscione, A. A. *et al.* CIAO: Chandra's data analysis system. *Proc. SPIE* **6270,** 1–12 (2006).
34. Arnaud, K. A. in *Astronomical Data Analysis Software and Systems V* (eds Jacoby, J. H. & Barnes, J.) 17–20 (ASP Conf. Ser. Vol. 101, 1996).
35. Verner, D. A., Ferland, G. J., Korista, K. T. & Yakovlev, D. G. Atomic data for astrophysics II. New analytic FITS for photoionization cross sections of atoms and ions. *Astrophys. J.* **465,** 487–510 (1996).
36. Wilms, J., Allen, A. & McCray, R. On the absorption of X-rays in the interstellar medium. *Astrophys. J.* **542,** 914–924 (2000).
37. Lu, F. J., Yuan, T. T. & Lou, Y.-Q. An imaging and spectral study of 10 X-ray filaments around the Galactic Center. *Astrophys. J.* **673,** 915–927 (2008).
38. Bamba, A. *et al.* X-ray evolution of pulsar wind nebulae. *Astrophys. J.* **719,** L116 (2010).
39. Quataert, E. & Loeb, A. Nonthermal THz to TeV emission from stellar wind shocks in the Galactic Center. *Astrophys. J.* **635,** L45–L48 (2005).
40. Cirelli, M., Kadastik, M., Raidal, M. & Strumia, A. Model-independent implications of the $e^\pm$, $\bar{p}$ cosmic ray spectra on properties of Dark Matter. *Nucl. Phys. B* **813,** 1–21 (2009).
41. Fan, J., Katz, A., Randall, L. & Reece, M. Double-disk dark matter. *Phys. Dark Univ.* **2,** 139–156 (2013).
42. Homan, J. *et al.* The X-ray properties of the black hole transient MAXI J1659-152 in quiescence. *Astrophys. J.* **775,** 9 (2013).
43. Degenaar, N. & Wijnands, R. A four-year baseline Swift study of enigmatic X-ray transients located near the Galactic Center. *Astron. Astrophys.* **524,** A69 (2010).
44. Zavlin, V. E. XMM-Newton observations of four millisecond pulsars. *Astrophys. J.* **638,** 951–962 (2006).
45. Manchester, R. N., Hobbs, G. B., Teoh, A. & Hobbs, M. The Australia Telescope National Facility pulsar catalogue. *Astron. J.* **129,** 1993–2006 (2005).
46. Takata, J., Cheng, K. S. & Taam, R. E. X-ray and gamma-ray emissions from rotation powered millisecond pulsars. *Astrophys. J.* **745,** 100 (2012).
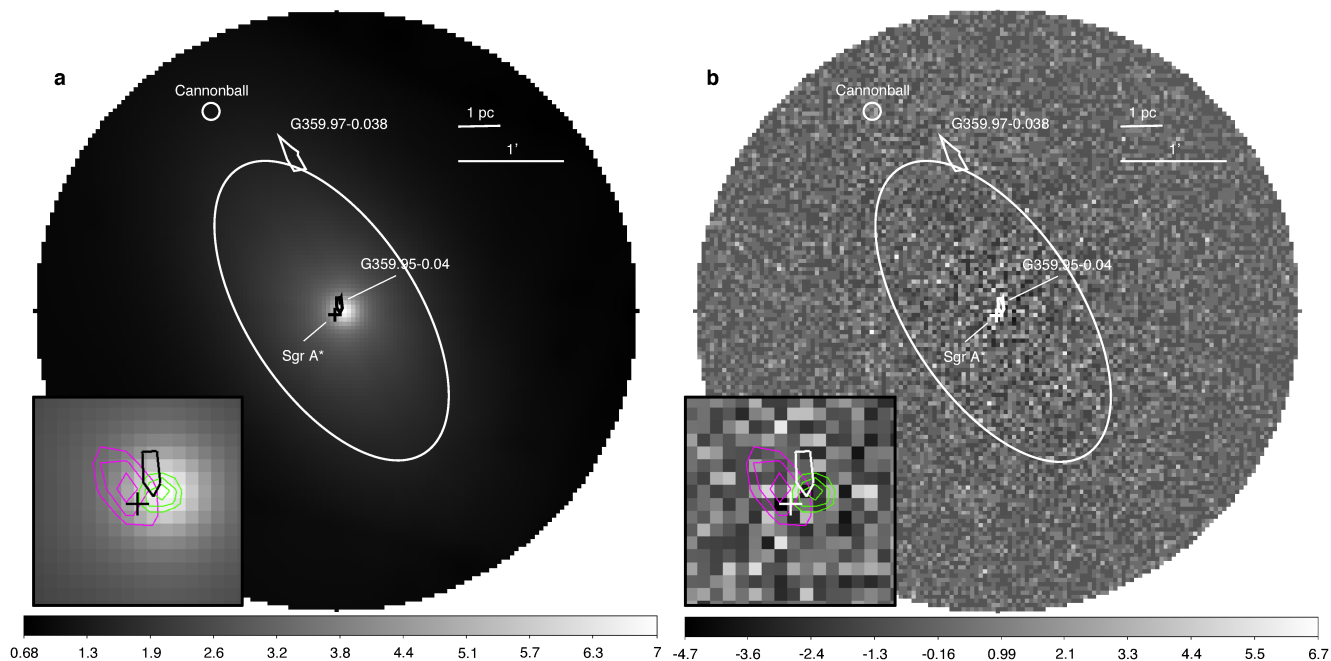
Extended Data Figure 1 | The best-fit 20–40 keV spatial model and residual image. a, The model, consisting of a point-like Gaussian, an extended Gaussian, and a background model as described in the main text. The radio position of Sgr A*[16] and the Chandra positions of the Cannonball[19], G359.95−0.04[13] and G359.97−0.038[22] are overlaid. The FWHM extent of the two-dimensional Gaussian fit to the extended emission is indicated by the solid ellipse. The image is shown with a linear grey-scale distribution with the range chosen to highlight the extended emission, in units of total counts per pixel (bar at bottom). **b**, The residual image (observed minus model). Insets in **a** and **b** show a magnified view of the central $40'' \times 40''$ of the main panels, with 1σ, 2σ and 3σ error contours on the centroid positions of the extended (magenta) and point-source (green) two-dimensional Gaussian models, as well as the position of Sgr A* (cross) and G359.95−0.04 (polygon).
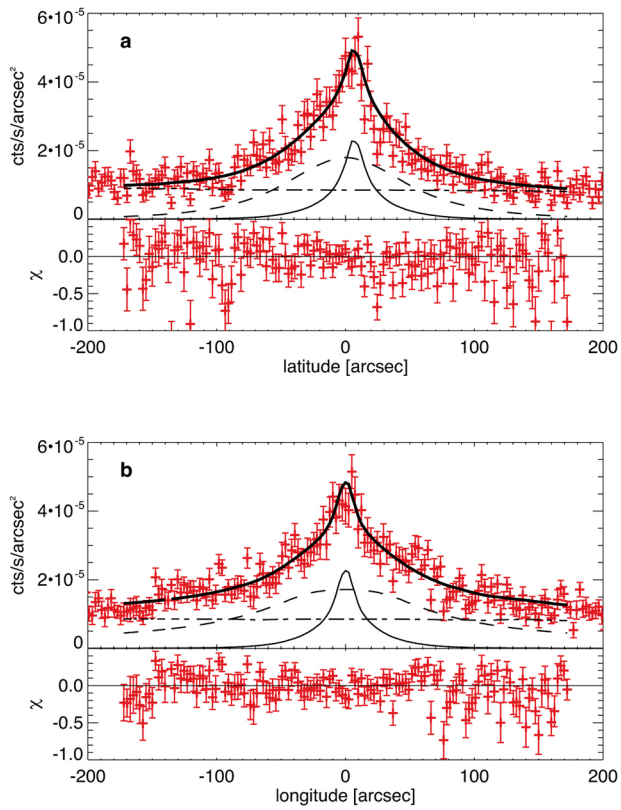
Extended Data Figure 2 │ Profiles of the 20–40 keV data and exponential spatial model along Galactic longitude and latitude. a, Galactic latitude; **b,** Galactic longitude. Top panels, the data (red, 1σ errors) are fitted to a two-dimensional model consisting of the surface brightness model used in ref. 5 and described in the main text (dashed line) and a point-like Gaussian source (thin solid line) that describes the emission from G359.95–0.04, both convolved with the on-axis NuSTAR PSF, as well as a flat background (dash-dot line) that varies between detector chips. Profiles are obtained by integrating from 25″ on either side of each axis, with the origin defined at the position of Sgr A*. The combined model (thick solid line) and residual emission (lower panels) show that this phenomenological model describes the data well.
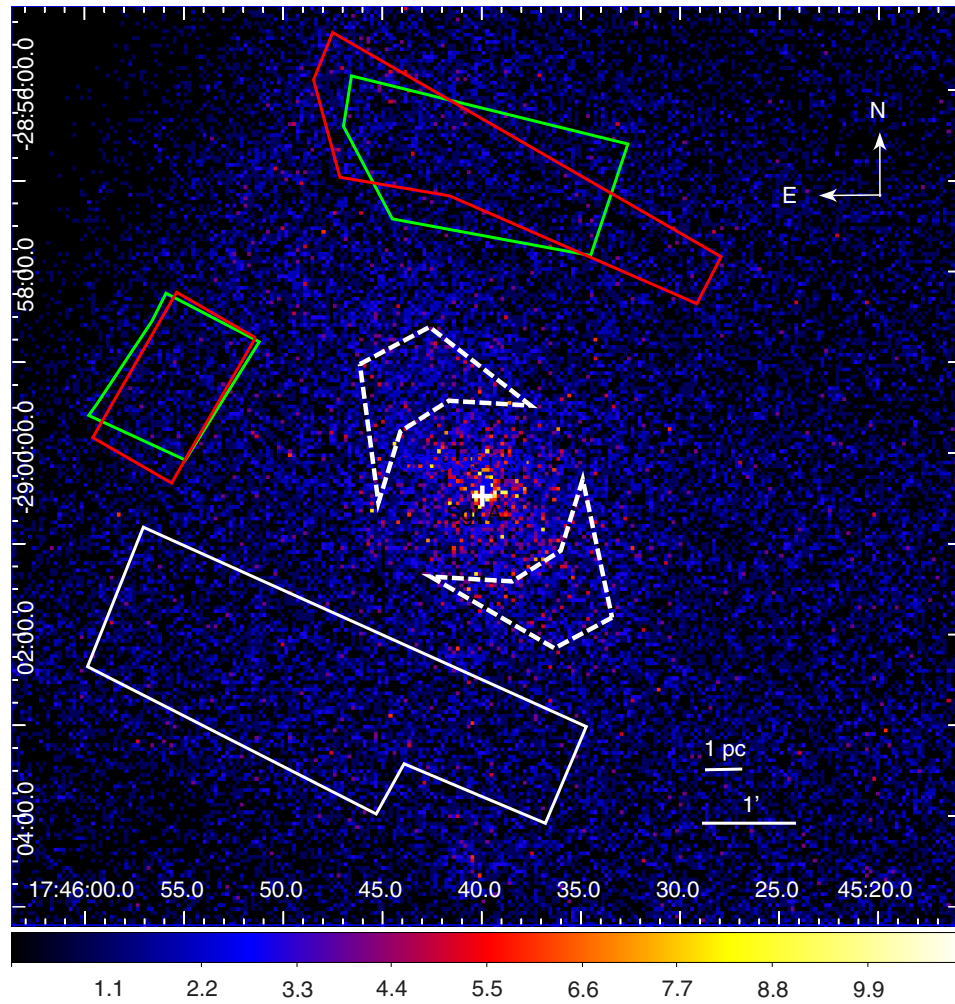
Extended Data Figure 3 | Regions used for extraction of background spectra for NuSTAR observations 30001002001 (red), 30001002003 (white), and 30001002004 (green). The regions used for spectral analysis of the diffuse emission are indicated by the dashed lines. These regions are overlaid on the NuSTAR 20–40 keV image. The colour scale shows units of total counts per pixel. The *x* and *y* axes indicate coordinates of right ascension and declination, respectively.
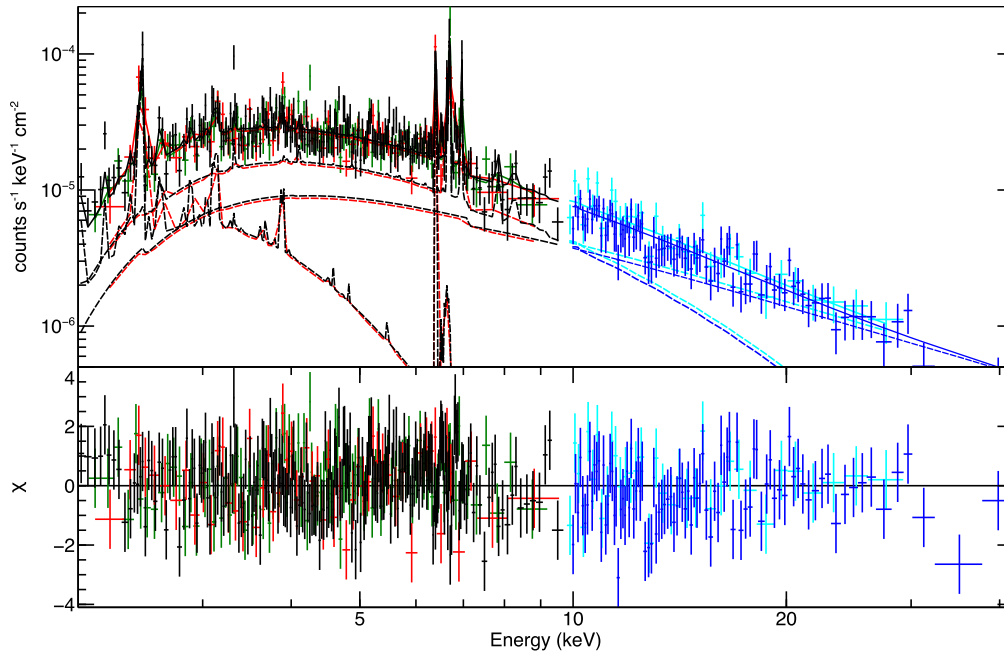
Extended Data Figure 4 | Unfolded 2–40 keV energy spectrum from the southwest region indicated in Fig. 1. Top panel, the 2–10 keV spectrum is constructed from XMM-Newton data from the PN (black), MOS1 (red), and MOS2 (green) instruments. The 10–40 keV spectrum is constructed from NuSTAR focal plane A (dark blue) and focal plane B (cyan) data. The fit to the data comprises two absorbed thermal plasmas plus an absorbed power-law, multiplied by a separate normalization factor for each data set to account for small calibration differences between instruments. The dashed lines indicate the separate model components. Fit residuals are shown in the lower panel. Full parameters of the spectral model are given in Extended Data Table 2. Error bars, 1σ statistical errors.
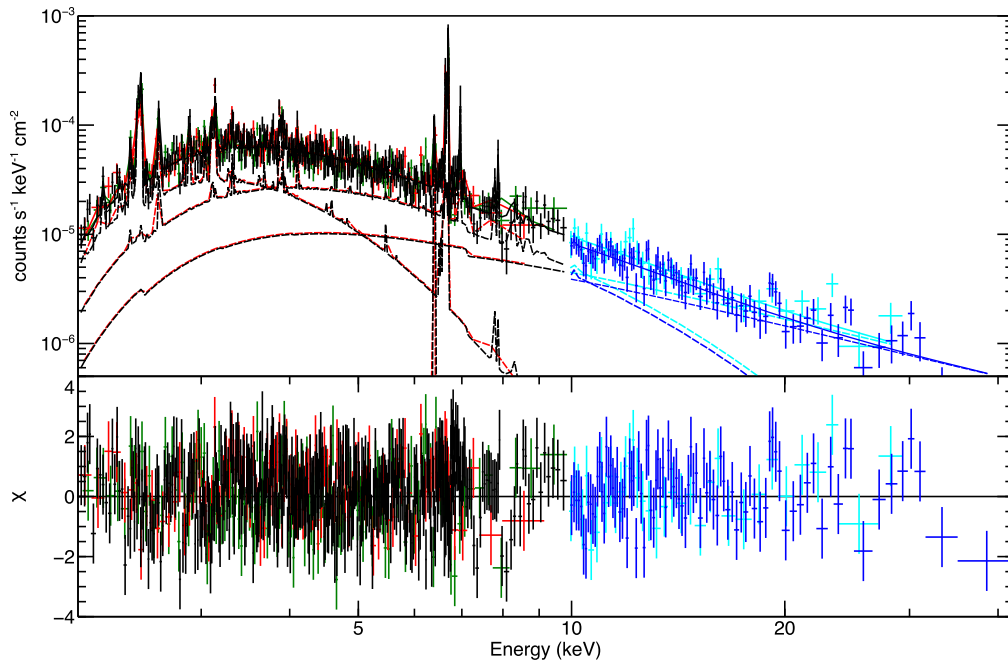
Extended Data Figure 5 | Unfolded 2–40 keV energy spectrum from the northeast region indicated in Fig. 1. Top panel, the spectrum is composed as in Extended Data Fig. 4. The model fitted to the data comprises an absorbed two-temperature thermal plasma plus a power-law, multiplied by a separate normalization factor for each data set. Fit residuals are shown in the lower panel. Full parameters of the spectral model are given in Extended Data Table 2. Error bars, 1σ statistical errors.
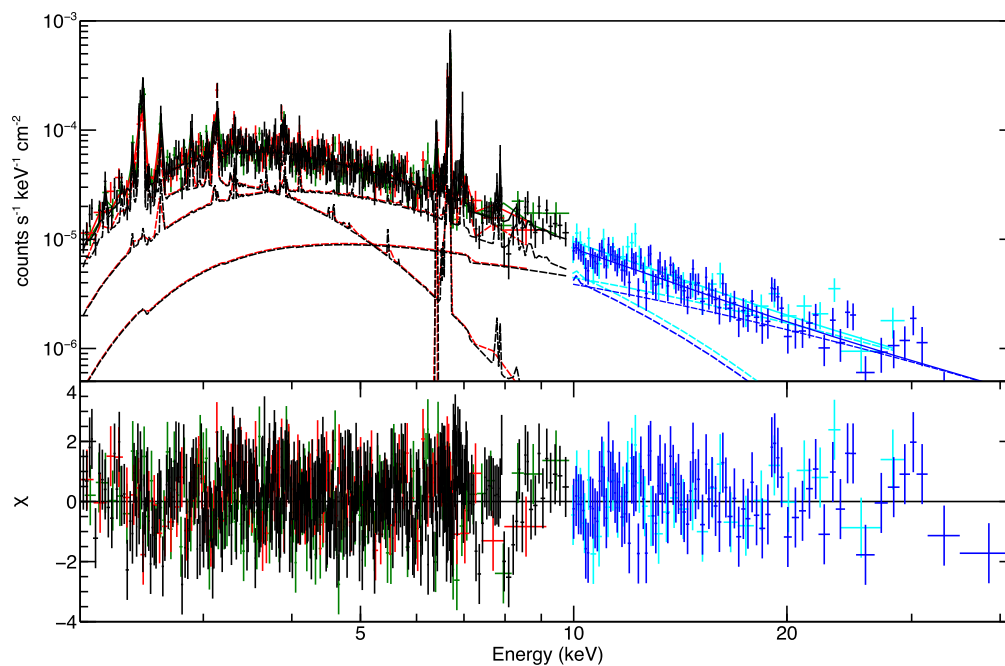
Extended Data Figure 6 | Unfolded 2–40 keV energy spectrum from the northeast region indicated in Fig. 1. The spectrum is composed as in Extended Data Fig. 4. The model fitted to the data comprises an absorbed two-temperature thermal plasma plus a thermal bremsstrahlung, multiplied by a separate normalization factor for each data set. Fit residuals are shown in the lower panel. Full parameters of the spectral model are given in Extended Data Table 3. Error bars, 1σ statistical errors.

Extended Data Table 1 | Best fit 2D Gaussian models

Parameter	G359.95-0.04	CHXE
R.A. (Center, J2000)	$266.4150^{\circ}{}^{+4.5"}_{-3.8"}$	$266.4172^{\circ}{}^{+6.6"}_{-6.2"}$
DEC (Center, J2000)	$-29.007245^{\circ}{}^{+3.4"}_{-4.1"}$	$-29.00716^{\circ}{}^{+8.1"}_{-7.1"}$
FWHM [arcseconds]	$1.8^{+2.3}_{-0.7}$	$195.8^{+21.9}_{-16.8}$
Amplitude [10^{-3} cts s^{-1}]	$5.2^{+7.1}_{-3.5}$	$0.0064^{+0.0012}_{-0.0017}$
Ellipticity	—	0.52
θ^a [degree]	—	57

Parameters of the best-fit two-dimensional (2D) Gaussian models describing the extended hard-X-ray emission and the central point-like source, which is spectrally and spatially consistent with the pulsar wind nebula G359.95−0.04. CHXE is the 20–40 keV extended central hard X-ray emission. The errors quoted are 3σ statistical errors. All parameters refer to source models before convolution with the NuSTAR PSF.
^{a}The angle θ is defined with respect to the positive northern axis.

Extended Data Table 2 | Fits with hard emission modelled as power law

Parameter	Southwest	Northeast
PN norm.[a]	$1.1^{+0.1}_{-0.1}$	$1.1^{+0.1}_{-0.1}$
MOS1 and MOS2 norm.	$1.0^{+0.1}_{-0.1}$	$1.1^{+0.1}_{-0.1}$
NuSTAR FPMA norm. (fixed)	1.0	1.0
NuSTAR FPMB norm.	$1.2^{+0.1}_{-0.1}$	$1.1^{+0.1}_{-0.1}$
N_H [10^{22} cm^{-2}]	$14.1^{+1.5}_{-1.3}$	$16.4^{+1.2}_{-0.8}$
Γ	$1.5^{+0.3}_{-0.2}$	$1.6^{+0.3}_{-0.4}$
N_Γ [10^{-4} photons cm^{-2} s^{-1} keV^{-1}]	$1.3^{+1.6}_{-0.7}$	$1.6^{+2.4}_{-1.1}$
kT_1 [keV]	$1.0^{+0.3}_{-0.4}$	$1.1^{+0.1}_{-0.2}$
N_{kT_1} [10^{-4} photons cm^{-2} s^{-1} keV^{-1}]	$10.8^{+103}_{-5.0}$	$89.9^{+56.1}_{-28.7}$
$Z_1[Z_\odot]^b$	$5.0^{+-}_{-3.6}$	$2.3^{+0.9}_{-0.4}$
kT_2 [keV]	$7.5^{+1.6}_{-1.3}$	$5.1^{+0.9}_{-0.7}$
N_{kT_2} [10^{-4} photons cm^{-2} s^{-1} keV^{-1}]	$9.2^{+1.7}_{-1.6}$	$17.5^{+6.7}_{-5.9}$
$Z_2[Z_\odot]^c$	1.7	$2.3^{+0.9}_{-0.4}$
Fe K-α eq. width [eV]	128^{+40}_{-31}	47^{+84}_{-16}
χ^2/d.o.f.	1.00 (503.4/503)	1.05 (807.1/770)
F_X(20-40 keV) [10^{-13} ergs cm^{-2} s^{-1}]^d	7.6	8.0

Spectral model of the two extended emission regions in the energy range 2–40 keV, obtained from a joint fit of XMM (2–10 keV) and NuSTAR (10–40 keV) data, with the hard emission modelled as a power law. All quoted errors are at 90% confidence level (CL). N_H, absorption column; N_Γ, N_{kT1}, N_{kT2} are the normalizations for the power-law, kT_1, and kT_2 components, respectively. [a]Relative normalizations between different instruments, defined with respect to NuSTAR FPMA. [b]Abundance relative to solar. These are best-fit values. In the southwest, Z_2 was then fixed during error calculations, as described in the text. [c]Metallicities are independent (linked) for the two temperature components in the southwest (northeast). [d]Observed flux.

Extended Data Table 3 | Fits with hard emission modelled as thermal bremsstrahlung

Parameter	Southwest	Northeast
PN norm.[a]	$1.1^{+0.1}_{-0.1}$	$1.2^{+0.1}_{-0.1}$
MOS1 and MOS2 norm.	$1.1^{+0.1}_{-0.1}$	$1.2^{+0.1}_{-0.1}$
NuSTAR FPMA norm. (fixed)	1.0	1.0
NuSTAR FPMB norm.	$1.2^{+0.1}_{-0.1}$	$1.2^{+0.1}_{-0.1}$
N_H [10^{22} cm^{-2}]	$13.4^{+1.6}_{-1.3}$	$16.4^{+1.2}_{-0.8}$
kT_{bremss} [keV]	58^{+127}_{-23}	66^{+203}_{-30}
N_{bremss} [10^{-4} photons cm^{-2} s^{-1} keV^{-1}]	$1.8^{+0.4}_{-0.4}$	$1.9^{+0.4}_{-0.3}$
kT_1 [keV]	$1.0^{+0.3}_{-0.3}$	$1.1^{+0.1}_{-0.2}$
N_{kT_1} [10^{-4} photons cm^{-2} s^{-1} keV^{-1}]	$9.5^{+30}_{-2.0}$	$93.2^{+55.0}_{-22.2}$
$Z_1[Z_\odot]^b$	$5.0^{+\ \ }_{-3.2}$	$2.2^{+0.4}_{-0.3}$
kT_2 [keV]	$7.2^{+1.4}_{-1.3}$	$5.0^{+0.9}_{-0.7}$
N_{kT_2} [10^{-4} photons cm^{-2} s^{-1} keV^{-1}]	$8.8^{+2.0}_{-1.8}$	$18.0^{+6.1}_{-4.8}$
$Z_2[Z_\odot]^c$	1.6	$2.2^{+0.4}_{-0.3}$
Fe K-α eq. width [eV]	123^{+90}_{-46}	46^{+12}_{-13}
χ^2/d.o.f.	1.00 (501.6/503)	1.05 (807.1/770)
F_X(20-40 keV) [10^{-13} ergs cm^{-2} s^{-1}]^d	7.3	8.0

Model and footnotes as in Extended Data Table 2, but with the hard emission modelled as a thermal bremsstrahlung. N_{bremss} is the normalization of the high-temperature bremsstrahlung component.

LETTER

Topological valley transport at bilayer graphene domain walls

Long Ju[1]*, Zhiwen Shi[1]*, Nityan Nair[1], Yinchuan Lv[1], Chenhao Jin[1], Jairo Velasco Jr[1], Claudia Ojeda-Aristizabal[1], Hans A. Bechtel[2], Michael C. Martin[2], Alex Zettl[1,3,4], James Analytis[1,3,4] & Feng Wang[1,3,4]

Electron valley, a degree of freedom that is analogous to spin, can lead to novel topological phases in bilayer graphene. A tunable bandgap can be induced in bilayer graphene by an external electric field[1-5], and such gapped bilayer graphene is predicted to be a topological insulating phase protected by no-valley mixing symmetry, featuring quantum valley Hall effects and chiral edge states[6-9]. Observation of such chiral edge states, however, is challenging because inter-valley scattering is induced by atomic-scale defects at real bilayer graphene edges[10]. Recent theoretical work[11-13] has shown that domain walls between AB- and BA-stacked bilayer graphene can support protected chiral edge states of quantum valley Hall insulators. Here we report an experimental observation of ballistic (that is, with no scattering of electrons) conducting channels at bilayer graphene domain walls. We employ near-field infrared nanometre-scale microscopy (nanoscopy)[14-16] to image _in situ_ bilayer graphene layer-stacking domain walls on device substrates, and we fabricate dual-gated field effect transistors based on the domain walls. Unlike single-domain bilayer graphene, which shows gapped insulating behaviour under a vertical electrical field, bilayer graphene domain walls feature one-dimensional valley-polarized conducting channels with a ballistic length of about 400 nanometres at 4 kelvin. Such topologically protected one-dimensional chiral states at bilayer graphene domain walls open up opportunities for exploring unique topological phases and valley physics in graphene.

Hexagonal two-dimensional crystals, such as graphene and transition-metal dichalcogenides, exhibit a pair of degenerate bands at the K and K′ valleys in momentum space. The valley electrons are characterized by non-trivial Berry curvatures, which give rise to anomalous quantum Hall states in graphene[17] as well as valley Hall effects (the valley counterpart of spin Hall effects[18,19]). Bilayer graphene provides an attractive platform from which to explore topological valley physics[20]: a tunable semiconductor bandgap can be induced in bilayer graphene using a vertical electrical field, and different topological phases can exist in gapped bilayer graphene[6,7,11,21-23]. In particular, quantum valley Hall insulator states can exist in bilayer graphene (with any edge termination other than the armchair type) with a field-induced bandgap, where counterpropagating chiral electrons with opposite valley index exist at their boundaries[6-9]. Such one-dimensional chiral states are topologically protected as long as there are no inter-valley scatterings generated by atomic-scale defects.

Domain walls between AB- and BA-stacked bilayer graphene provide an attractive place to realize one-dimensional chiral boundary states of quantum valley Hall insulators because a smooth domain wall preserves the electron valley index in bilayer graphene, unlike valley-mixing atomic defects at graphene edges. The integrated Berry curvature of bilayer graphene valence bands for each valley is characterized by a quantized "valley" Chern number[11-13], the sign of which depends both on the direction of the vertical electrical field and the layer

stacking order. Martin _et al._[11] first proposed topological domain walls between band-inverted bilayer graphene insulators generated by opposite vertical electrical fields. Experimental realization of such electric-field walls, however, is extremely challenging. An alternative approach is to exploit different stacking orders. AB- and BA-stacked bilayer graphene under the same vertical electrical field represent two distinct topological phases with opposite valley Chern numbers. The stacking boundary between AB and BA domains forms a layer stacking domain wall (Fig. 1a). Valley-polarized chiral electrons are predicted[12,13] to flow along such a domain wall in gapped bilayer graphene, where K and K′ valley electrons propagate in opposite directions. Recent transmission electron microscopy studies show that smooth AB-BA domain walls with a width of several nanometres exist naturally in suspended bilayer graphene[24,25], but such suspended transmission electron microscopy bilayer graphene samples on through-holes (that is, holes that penetrate through a substrate) cannot easily be combined with bandgap control and transport measurements. Here we use a new approach to image _in situ_ AB-BA domain walls in exfoliated (that is, prepared from graphite using tape) bilayer graphene on SiO_2/Si substrates, and demonstrate topologically protected charge and valley transport along such boundaries for the first time.

Few-layer graphene was mechanically exfoliated on 280-nm-thick SiO_2/Si substrates, and the layer thickness is determined using optical contrast measurements. To identify AB-BA layer stacking domain walls in bilayer graphene we used near-field infrared nanoscopy[14], as illustrated in Fig. 1b. Infrared light at 6.1 μm was focused onto the apex of a metal-coated atomic force microscope (AFM) tip with curvature radius $r \approx 25$ nm, and the scattered infrared radiation was collected and measured by an HgCdTe detector in the far field. The tip-enhanced infrared scattering provides a local probe of the material infrared responses with ~40 nm spatial resolution. The AB- and BA-stacked bilayer graphene, being inversion symmetric to each other, have exactly the same electronic bands and infrared absorption. The local band structure of the domain wall, however, is strongly modified relative to the bulk and exhibits different infrared contrast. Figure 1c and d displays AFM topography and near-field infrared nanoscopy images of the same bilayer graphene on a SiO_2/Si substrate. The bilayer graphene topography shows a step height of ~1 nm, consistent with the bilayer thickness. Within the bilayer graphene flake, the topography is featureless with a height fluctuation of less than 0.2 nm. In the near-field infrared nanoscopy image, however, distinct features appear in the bilayer graphene: multiple bright lines extend across the bilayer graphene flake and terminate at the edges. These bright lines in the infrared image indicate a change of local infrared conductivity in the bilayer graphene. No corresponding features can be discerned in the topography image (Fig. 1c), which excludes the possibility of surface contaminations or graphene folding along these lines. We attribute these local infrared response changes to AB-BA domain walls in bilayer graphene.

[1]Department of Physics, University of California, Berkeley, California 94720, USA. [2]Advanced Light Source Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [3]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]Kavli Energy NanoSciences Institute at the University of California, Berkeley and the Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
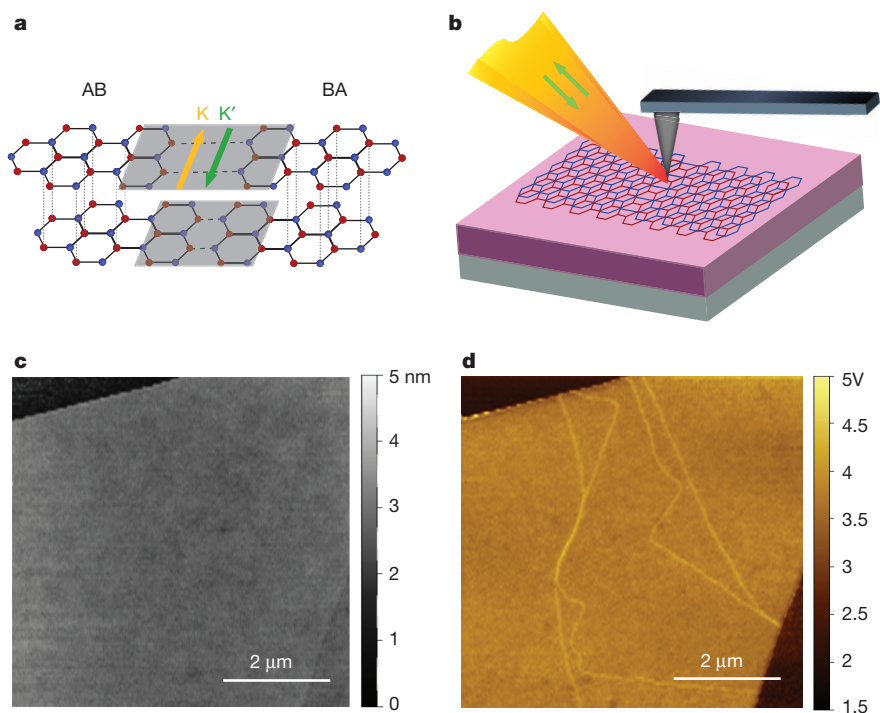*These authors contributed equally to this work.

a

b

c

d

Figure 1 | Imaging AB-BA domain walls in exfoliated bilayer graphene. **a**, Illustration of a domain wall (shaded area) between AB- and BA-stacked bilayer graphene domains. The yellow (green) arrow indicates the one-dimensional conducting channel at the domain wall from K-valley (K′-valley) electrons. **b**, Illustration of the near-field infrared nanoscopy measurement of graphene on a SiO_2/Si substrate (pink/grey). Infrared light (yellow) at 6.1 μm was focused onto the apex of a metal-coated AFM tip and the backscattered infrared radiation was collected and measured by an HgCdTe detector in the far

field. **c**, AFM topography map of a bilayer graphene sample on a SiO_2/Si substrate, showing no surface wrinkles or defects. (The small triangular area at the bottom right corner corresponds to a graphene monolayer.) **d**, Near-field infrared nanoscopy image of the same bilayer graphene sample as in **c**. Bright line features across the bilayer graphene flakes are observed that are absent in the topography image. These bright lines arise from AB-BA domain walls in bilayer graphene because the domain walls have different local electronic structure and infrared responses.

Qualitatively, we can approximate bilayer graphene in the domain wall region as randomly stacked bilayers, which has stronger local infrared response than bulk AB-stacked bilayer graphene and leads to the bright lines observed in infrared nanoscopy images (see Methods for details).

To confirm that the bright lines within the bilayer graphene are layer stacking domain walls, we systematically examined near-field infrared images of exfoliated graphene flakes of different layer thicknesses. Figure 2a–c shows infrared nanoscopy images of representative mono-layer, bilayer, and trilayer graphene flakes, respectively. The infrared images of monolayers, like that in Fig. 2a, always show homogenous responses within the flake, although bright features can be observed parallel to the edges owing to the excitation of graphene plasmons. The absence of any structural domain boundary in monolayers is consistent with the high quality of the exfoliated samples. In bilayer graphene, about 30% of the exfoliated flakes show bright lines across the interior region of graphene like those in Fig. 2b, and areas between these lines have the same contrast. We identify these bright lines as AB-BA domain walls because they are the only one-dimensional structure that exists in bilayer but not monolayer graphene. This identification is further bolstered by the infrared images of graphene trilayers, such as that shown in Fig. 2c.

Bright lines extending through the graphene flakes are also observed, but unlike graphene bilayers, the domains separated by the bright lines can have different infrared contrast. This is because more domain stacking orders can exist in the trilayer, such as ABA, BAB, or ABC stacking. Although ABA and BAB domains are inversion symmetric to each other and have the same electronic structure, ABC domains have distinct band structures[26], infrared responses, and Raman spectra[27]. Therefore, a domain wall between ABA- and ABC-stacked trilayers will separate two domains having different infrared

contrast in the near-field infrared image, as seen in domains I and II in Fig. 2c. In addition, we performed micro-Raman measurements on domains I and II, and show the corresponding spectra in Fig. 2d. It is apparent that the two-dimensional Raman features in domains I and II are different, and they have been shown to correspond to ABC- and ABA-stacked trilayer graphene[27], respectively. It demonstrates unambiguously that the bright line features in trilayers are layer stacking domain walls, and supports our assignment for bilayers.

After identifying the layer stacking domain walls in bilayer graphene on a SiO_2/Si substrate, we fabricated dual-gated bilayer field-effect transistor devices[4,5] across the domain walls. As references, we also fabricated dual-gated devices without domain walls using the same device configuration on the same bilayer graphene flakes. Figure 3a displays the optical microscopy image of a representative device. The white dashed line outlines the bilayer graphene flake, and the green line denotes the position of the AB-BA domain wall in the bilayer graphene. Source and drain electrodes (electrodes 1–5) composed of 40 nm Au and 0.5 nm Cr layers were defined by electron-beam lithography. Electrodes 1 and 2 are connected by a layer stacking domain wall, as are electrodes 2 and 3. Electrodes 3 and 4 and electrodes 4 and 5, on the other hand, are within a single bilayer graphene domain and act as reference devices. On top of the bilayer graphene flake, a 35-nm-thick Al_2O_3 dielectric layer was grown using atomic layer deposition. Last, the top gate electrodes of width 1 μm (TG1 and TG2) were patterned using electron-beam lithography and metal deposition. Figure 3b illustrates the side view of the device between electrodes 2 and 4, with a domain wall (green line) between electrodes 2 and 3 and single-domain bilayer graphene between electrodes 3 and 4. Here the doped silicon substrate acts as a bottom gate. By varying the top gate voltage V_{tg} and the bottom gate voltage V_{bg}, we can independently control the electric field above and below the bilayer graphene to generate a tunable semiconductor bandgap[5].
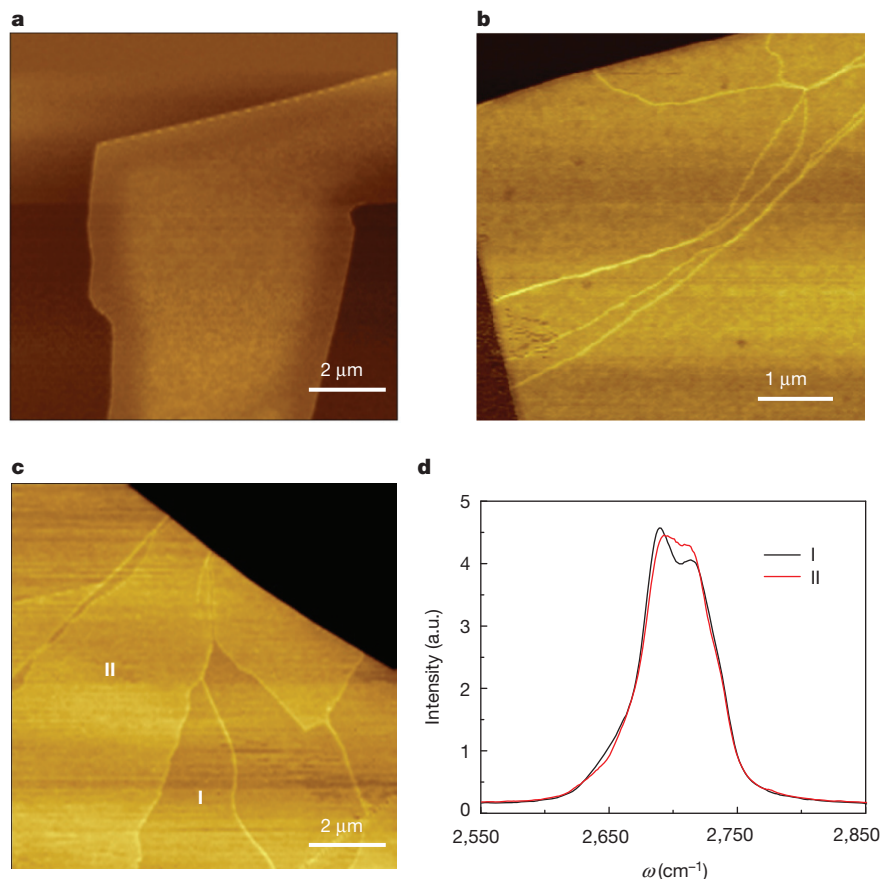
Figure 2 | Infrared nanoscopy images of domains and domain walls in exfoliated mono-, bi- and trilayer graphene. a, Infrared image of a graphene monolayer. The darker region corresponds to the SiO_2/Si substrate. Monolayer graphene shows homogeneous contrast across the whole sample with no bright lines in the interior area. (The bright lines parallel to the edge arise from plasmon excitations.) The same behaviour is observed in all exfoliated monolayers. **b**, Infrared image of a graphene bilayer, which shows bright line features across the flake. Such bright lines arise from AB-BA domain walls in bilayer graphene and they are observed in $\sim$30% of exfoliated bilayer graphene samples. The bulk AB and BA domains have the same electronic structure and infrared contrast. **c**, Infrared image of a graphene trilayer, which exhibits domains with different infrared contrast as well as bright lines between domains. Different stacking orders, such as ABA and ABC stacking, can lead to different electronic structures and infrared contrast in trilayers. **d**, Raman spectra taken at domains I and II in **c**; ω is the Raman shift. The two-dimensional peak of domain I (black curve) and domain II (red curve) are characteristic of ABC and ABA stacking, respectively. It confirms that the bright and dark trilayer regions in **c** correspond to different layer stacking orders, and that the bright line features arise from layer stacking domain walls.

Figure 3c displays the electrical transport at 4 K in a reference bilayer graphene device without a domain wall. Each trace shows the gate-dependent resistance as a function of V_{tg} at a fixed bottom gate voltage, V_{bg}. V_{bg} was varied in steps of 10 V to produce the different traces. The resistance peak in each trace corresponds to a charge neutral point (CNP), at which the Fermi level resides within the semiconductor bandgap. CNPs can be realized with different combinations of V_{tg} and V_{bg}, but each with a different vertical electrical field and induced semiconductor bandgap in bilayer graphene. The zero-bandgap state is realized at $V_{bg}^0 = 0\,V$ and $V_{tg}^0 = 1.8\,V$, which has the lowest peak resistance. The average displacement field[5] across bilayer graphene is described by $D = \frac{1}{2}(D_b + D_t)$. Here $D_b = \varepsilon_{SiO_2}|V_{bg} - V_{bg}^0|/d_{SiO_2}$, and $D_t = \varepsilon_{Al_2O_3}|V_{tg} - V_{tg}^0|/d_{Al_2O_3}$, where ε and d are the dielectric constant and thickness of the respective oxide dielectric layers. The bandgap increases monotonically with the average vertical displacement field D, which results in a larger peak resistance at the CNP. The experimental values of the peak resistances are lower than that expected from the semiconductor bandgap in ideal bilayer graphene, presumably owing to impurities and defects in our devices. Nevertheless, the resistance can reach $\sim$90 kΩ at 4.2 K, similar to that observed in previous dual-gated bilayer graphene devices on SiO_2/Si substrates[4].

Electrical transport through bilayer graphene with an AB-BA domain wall exhibits a different behaviour, as shown in Fig. 3d. When the bilayer graphene bandgap is small (purple traces), the gate-dependent resistance curves are similar to those in single-domain bilayer graphene devices because bulk current flowing through the domain dominates. With higher displacement fields and larger bilayer graphene bandgaps, the peak resistance in each curve first increases and then saturates at $\sim$14 kΩ (red traces in Fig. 3d), in striking contrast to the monotonically increasing resistance to over 80 kΩ in the reference device without a domain wall (Fig. 3c). Figure 4a plots the comparison of bilayer graphene peak resistances in the device with a domain wall (red circles) and without a domain wall (black squares) as a function of the vertical displacement field D, explicitly showing that saturation of the peak resistances occurs at a much lower value, owing to the presence of the domain wall.

These distinct transport behaviours in bilayer graphene with and without AB-BA domain walls were observed in all of the bilayer graphene devices that we studied (see Methods for details.) It is apparent that the conductance is always much higher in bilayer graphene devices with AB-BA domain walls. In addition, we performed temperature dependence studies of CNP resistance in bilayer graphene with and without domain walls from 1.8 K to 50 K. We observed that the resistance of bulk bilayer graphene increases many times at lower temperature, characteristic of an insulating phase. The resistance across the domain wall, however, behaves like a metallic system with much weaker temperature dependence (see Methods for details). Our results demonstrate unambiguously that one-dimensional conducting
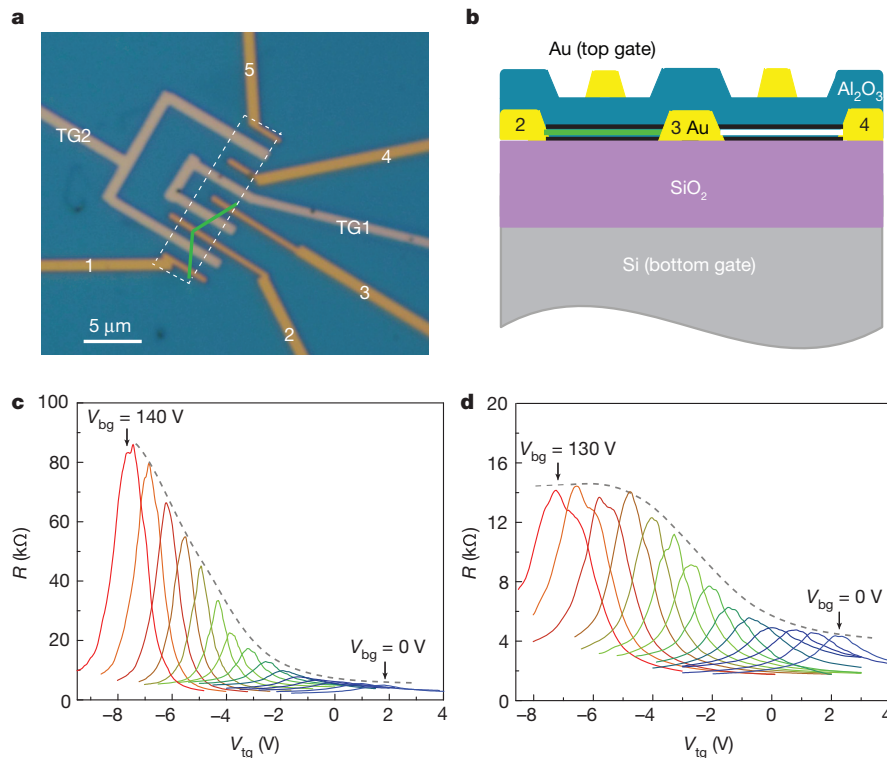
Figure 3 | Dual-gated field-effect transistor devices on bilayer graphene with AB-BA domain walls and their electrical transport properties.
a, Optical micrograph of a dual-gated bilayer device. The white dotted line outlines the bilayer graphene flake, and the green line denotes the AB-BA domain wall. Source and drain electrodes are labelled from 1 to 5. Electrodes 1 and 2 are connected by a layer stacking domain wall, as are electrodes 2 and 3. Electrodes 3 and 4 and electrodes 4 and 5, on the other hand, are within a single bilayer graphene domain and act as reference devices. TG1 and TG2 denote top gate electrodes. **b**, Side view of the dual-gated bilayer device between electrodes 2 and 4 in **a**. **c**, Gate-dependent resistance for bilayer graphene without a

domain boundary (electrodes 3 and 4). The backgate voltage V_{bg} is varied from 0 V to 140 V with a step of 10 V (colours here and in **d** indicate different curves), and the top gate voltage V_{tg} is swept continuously. The resistance peak in each trace corresponds to a CNP. The resistance at the CNP increases continuously (dashed grey line) to over 80 kΩ at high V_{bg} owing to the opening of a bandgap. **d**, Gate-dependent resistance across bilayer graphene with a domain boundary (electrodes 2 and 3). The resistance at the CNP first increases and then saturates at ~14 kΩ, although the bilayer graphene bandgap keeps increasing with increasing V_{bg}. This much lower resistance signifies the presence of a one-dimensional conducting channel at the AB-BA domain wall.

channels are present at the AB-BA domain walls, and that they dominate the electrical transport when bulk bilayer graphene domains become insulating.

To examine the conducting channels through the bilayer graphene domain walls, we systematically investigated bilayer graphene devices with varying channel length (defined by the top gate width) along the same bilayer graphene domain wall. Figure 4b displays the measured conductance of the two sets of bilayer graphene domain wall channels with lengths of 200 nm, 400 nm, 600 nm and 800 nm, respectively and another batch of 1,000-nm-long bilayer graphene domain wall channels. All the conductance values are of the order of the quantum conductance ($2e^2/h$, where e is the unit charge and h is Planck's constant), and one of the shortest channels exhibits a conductance approaching $4e^2/h$.

The observed one-dimensional conducting channels correspond to the chiral edge modes at AB-BA domain walls recently predicted by theory[11–13]. Figure 4c shows the electronic band structure around the K and K' points for bilayer graphene with an AB-BA domain wall. In gapped bilayer graphene, the AB- and BA-stacked domains represent two different quantum valley Hall insulators. Consequently, a domain wall separating insulating AB and BA domains will confine a pair of one-dimensional conducting channels along the boundary for each valley. The conducting electrons in different valleys have opposite energy dispersions, and therefore propagate with opposite velocities. Such counter-propagating, valley-projected chiral electrons along the AB-BA domain walls are illustrated in Fig. 4d in real space. These conducting chiral modes are topologically protected within a single valley. Inter-valley scatterings can potentially mix the forward- and backward-moving electrons and eliminate the chiral modes, analogous to magnetic impurities

that flip spins and eliminate chiral modes in quantum spin Hall insulators. However, at atomic-defect-free AB-BA domain walls where the crystal lattice is smoothly varying, such inter-valley scattering will be absent because it requires a large momentum change of $4\pi/(3\sqrt{3})a_{c-c}$, where a_{c-c} is the carbon–carbon bond length of 1.42 Å.

We can quantitatively compare the observed one-dimensional conductance along the AB-BA domain wall with theoretical predictions. As illustrated in Fig. 4d, two conducting chiral modes are present at the domain wall for a given current flow direction. In addition, each chiral mode is doubly degenerate owing to electron spin. Therefore, four one-dimensional conducting channels exist at the domain wall, giving an ideal conductance of $4e^2/h$, that is, two conductance quanta. Experimentally, we observed conductance of the order of the conductance quanta in all devices, and one 200-nm-long domain wall channel exhibits conductance close to the value of $4e^2/h$ (Fig. 4b). We found that overall the one-dimensional channel conductance becomes lower with increasing channel length. The observed length-dependent conductance through bilayer graphene domain walls can be described by the Landauer–Büttiker formula[28] with a finite electron mean free path: $G = \left(R_c + R_0\left(1 + \frac{L}{L_0}\right)\right)^{-1}$, where L is the channel length, $R_0 = h/4e^2$, the contact resistance $R_c = 0$, and the mean free path $L_0 = 420$ nm (solid line in Fig. 4b). The electron mean free path of 420 nm can arise from a finite inter-valley scattering along the bilayer graphene domain walls, and previous transport studies have reported a comparable inter-valley scattering length of 300–800 nm in bilayer graphene on SiO$_2$/Si substrates[29]. Nevertheless, this electron mean free path along the bilayer graphene domain wall is orders of magnitude longer than
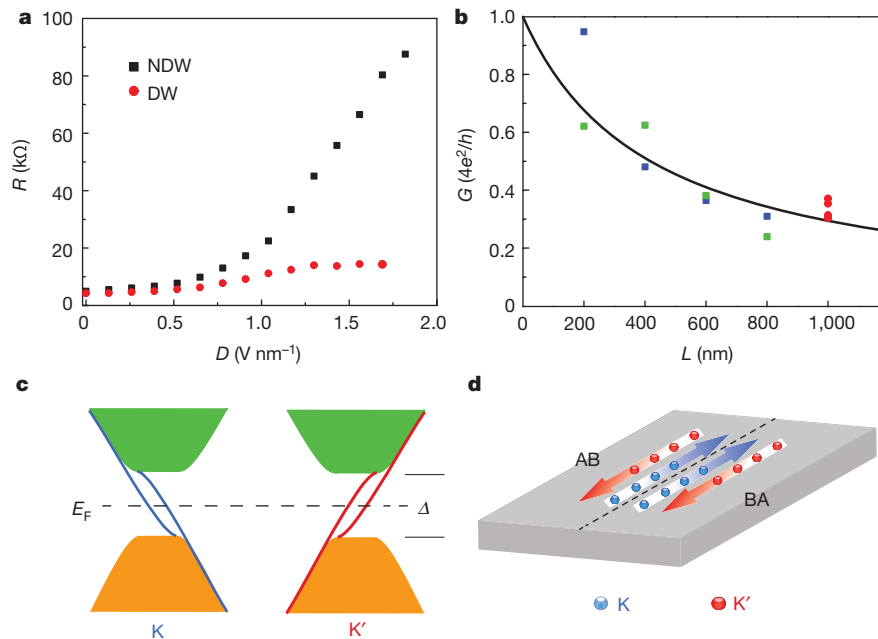
Figure 4 | Conductivity of topologically protected chiral modes at AB-BA domain walls. a, CNP resistances as a function of displacement field D for bilayer graphene with no domain wall (NDW, black squares) and for bilayer graphene with a domain wall (DW, red circles) obtained from Fig. 3. The saturation of the resistance at a much lower value at high D originates from one-dimensional conducting channels at the domain wall. **b,** Length dependence of the bilayer graphene domain wall conductance. Green and blue symbols represent two sets of devices, each fabricated on one single domain wall in the same bilayer graphene flake. Red symbols represent another batch of 1,000-nm-long domain wall channels. All devices exhibit conductance of the order of the quantum conductance, with lower conductance in longer channels and a maximum conductance approaching $4e^2/h$. This length-dependent one-dimensional conductance can be described by four conducting channels with a mean free path of 420 nm using the Landauer–Büttiker formula (solid line). **c,** Electronic band structure of bilayer graphene with an AB-BA domain wall. Solid green and orange regions represent states of the bulk bilayer with a bandgap of Δ, E_F indicates the Fermi energy, and the blue (red) lines represent topologically protected K-valley (K′-valley) chiral electron modes at the domain wall. Each chiral electron mode is doubly spin degenerate, leading to four valley-polarized one-dimensional conducting channels at the AB-BA domain wall of a gapped bilayer graphene sample. **d,** Illustration of valley-polarized electrical current at the bilayer graphene AB-BA domain wall. Boundary states in the K and K′ valley move in opposite one-way channels, leading to ballistic charge and valley current along the domain wall.

that in the bilayer graphene domain, which has a mobility of $\sim$2,000 cm^2 V^{-1} s^{-1} and a mean free path of only $\sim$8 nm in our samples. In general, scattering and localization effects tend to be more important in one dimension. The much longer mean free path at the AB-BA domain walls therefore highlights the topological protection of chiral electron modes, where the electron velocity-valley locking forbids intra-valley backscattering. Because inter-valley scattering along the smooth domain walls is negligible, the chiral valley channels at AB-BA domain walls are very robust even for bilayer graphene devices on SiO$_2$/Si substrates. We envision that encapsulation in boron nitride layers[30] could potentially lead to chiral electrons at AB-BA domain walls with very long coherent lengths.

Our observation of topologically protected one-dimensional conducting channels at bilayer graphene domain walls opens up exciting opportunities to explore tunable topological phases and valley physics in graphene. Owing to the velocity-valley locking, the one-dimensional chiral modes at the AB-BA domain walls lead not only to ballistic charge transport, but also to ballistic valley transport. Such chiral modes can enable novel ways to control the valley degree of freedom for valleytronic devices based on bilayer graphene.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. McCann, E. Asymmetry gap in the electronic band structure of bilayer graphene. *Phys. Rev. B* **74,** 161403 (2006).
2. Castro, E. V. *et al.* Biased bilayer graphene: semiconductor with a gap tunable by the electric field effect. *Phys. Rev. Lett.* **99,** 216802 (2007).
3. Min, H. K., Sahu, B., Banerjee, S. K. & MacDonald, A. H. *Ab initio* theory of gate induced gaps in graphene bilayers. *Phys. Rev. B* **75,** 155115 (2007).
4. Oostinga, J. B., Heersche, H. B., Liu, X. L., Morpurgo, A. F. & Vandersypen, L. M. K. Gate-induced insulating state in bilayer graphene devices. *Nature Mater.* **7,** 151–157 (2008).
5. Zhang, Y. B. *et al.* Direct observation of a widely tunable bandgap in bilayer graphene. *Nature* **459,** 820–823 (2009).
6. Yao, W., Yang, S. A. & Niu, Q. Edge states in graphene: from gapped flat-band to gapless chiral modes. *Phys. Rev. Lett.* **102,** 096801 (2009).
7. Zhang, F., Jung, J., Fiete, G. A., Niu, Q. A. & MacDonald, A. H. Spontaneous quantum Hall states in chirally stacked few-layer graphene systems. *Phys. Rev. Lett.* **106,** 156801 (2011).
8. Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81,** 109 (2009).
9. Jung, J., Zhang, F., Qiao, Z. H. & MacDonald, A. H. Valley-Hall kink and edge states in multilayer graphene. *Phys. Rev. B* **84,** 075418 (2011).
10. Yan, J. & Fuhrer, M. S. Charge transport in dual gated bilayer graphene with Corbino geometry. *Nano Lett.* **10,** 4521–4525 (2010).
11. Martin, I., Blanter, Y. M. & Morpurgo, A. F. Topological confinement in bilayer graphene. *Phys. Rev. Lett.* **100,** 036804 (2008).
12. Zhang, F., MacDonald, A. H. & Mele, E. J. Valley Chern numbers and boundary modes in gapped bilayer graphene. *Proc. Natl Acad. Sci. USA* **110,** 10546–10551 (2013).
13. Vaezi, A., Liang, Y. F., Ngai, D. H., Yang, L. & Kim, E. A. Topological edge states at a tilt boundary in gated multilayer graphene. *Phys. Rev. X* **3,** 021018 (2013).
14. Keilmann, F. & Hillenbrand, R. Near-field microscopy by elastic light scattering from a tip. *Phil. Trans. R. Soc. A* **362,** 787–805 (2004).
15. Chen, J. N. *et al.* Optical nano-imaging of gate-tunable graphene plasmons. *Nature* **487,** 77–81 (2012).
16. Fei, Z. *et al.* Gate-tuning of graphene plasmons revealed by infrared nano-imaging. *Nature* **487,** 82–85 (2012).
17. Novoselov, K. S. *et al.* Unconventional quantum Hall effect and Berry's phase of 2π in bilayer graphene. *Nature Phys.* **2,** 177–180 (2006).
18. Kane, C. L. & Mele, E. J. Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95,** 226801 (2005).
19. Konig, M. *et al.* Quantum spin hall insulator state in HgTe quantum wells. *Science* **318,** 766–770 (2007).
20. Wen, X. G. Symmetry-protected topological phases in noninteracting fermion systems. *Phys. Rev. B* **85,** 085103 (2012).

21. Weitz, R. T., Allen, M. T., Feldman, B. E., Martin, J. & Yacoby, A. Broken-symmetry states in doubly gated suspended bilayer graphene. *Science* **330,** 812–816 (2010).
22. Velasco, J. *et al.* Transport spectroscopy of symmetry-broken insulating states in bilayer graphene. *Nature Nanotechnol.* **7,** 156–160 (2012).
23. Maher, P. *et al.* Evidence for a spin phase transition at charge neutrality in bilayer graphene. *Nature Phys.* **9,** 154–158 (2013).
24. Alden, J. S. *et al.* Strain solitons and topological defects in bilayer graphene. *Proc. Natl Acad. Sci. USA* **110,** 11256–11260 (2013).
25. Butz, B. *et al.* Dislocations in bilayer graphene. *Nature* **505,** 533–537 (2013).
26. Min, H. K. & MacDonald, A. H. Electronic structure of multilayer graphene. *Prog. Theor. Phys.* **176** (Suppl.), 227–252 (2008).
27. Lui, C. H. *et al.* Imaging stacking order in few-layer graphene. *Nano Lett.* **11,** 164–169 (2011).
28. Datta, S. *Electronic Transport in Mesoscopic Systems* Ch. 2 (Cambridge Univ. Press, 1995).
29. Gorbachev, R. V., Tikhonenko, F. V., Mayorov, A. S., Horsell, D. W. & Savchenko, A. K. Weak localization in bilayer graphene. *Phys. Rev. Lett.* **98,** 176805 (2007).
30. Wang, L. *et al.* One-dimensional electrical contact to a two-dimensional material. *Science* **342,** 614–617 (2013).

METHODS

Mechanism of the infrared contrast for domain walls. Our near-field infrared nanoscopy probes the local optical conductivity at infrared frequencies with ~40 nm spatial resolution. The infrared contrast of the domain wall arises from its different electronic structure and infrared conductivity from the AB-stacked bilayer graphene. An exact description of the local band structure of a domain wall is beyond the scope of this work. However, the domain wall region is characterized by shifted top and bottom layers, which can be approximated by two randomly stacked monolayers of graphene. Such randomly stacked graphene has electronic structure similar to that of two separated monolayers close to the Dirac point[31], and its infrared conductivity can be approximated by $2\sigma_{monolayer}$. The AB-stacked bilayer graphene, on the other hand, has an infrared conductivity of $\sigma_{bilayer}$. The optical conductivities of both monolayer and AB-stacked bilayer graphene have been well established in the literature[32]. At the infrared wavelength of 6.1 μm, the value of $2\sigma_{monolayer}$ is about 20% higher than that of $\sigma_{bilayer}$, which can qualitatively explain the bright infrared contrast of the domain walls observed in our studies.

Electrical transport data of four extra bilayer graphene devices. We investigated five bilayer graphene device pairs with a fixed channel length of 1 μm. In Extended Data Fig. 1, we show the complete set of data taken from the four pairs of bilayer graphene devices other than that included in Fig. 3c and d. Each pair of bilayer graphene devices includes one with AB-BA stacking domain walls and one reference device with no domain walls on the same graphene flake. Extended Data Fig. 1a–d displays resistances of corresponding bilayer graphene devices with domain walls, including three single-domain-wall devices (Extended Data Fig. 1a–c) and one double-domain-wall device (Extended Data Fig. 1d). All single-domain-wall devices show resistances saturating at 14 kΩ with increasing D, and the double-domain-wall device has a resistance saturating around 8 kΩ. For comparison, Extended Data Fig. 1e–h displays the gate-dependent resistance for four bilayer graphene devices without boundaries. These devices show continuously increasing CNP resistance to well above 14 kΩ (8 kΩ in Extended Data Fig. 1h), as indicated by the dashed line at high V_{bg}. All the data show that one-dimensional conducting channels are present at AB-BA domain walls, and that they dominate electrical transport in gapped bilayer graphene.

Length dependence of electrical transport along AB-BA domain walls. We investigated two sets of dual-gated bilayer graphene domain wall devices of varying channel lengths (defined by the top gate width). Each set of devices are fabricated on one single domain wall in the same bilayer graphene flake, and includes four channels with lengths of 200 nm, 400 nm, 600 nm and 800 nm, respectively. The complete transport data of these bilayer graphene domain wall devices are shown in Extended Data Fig. 2.

Temperature dependence of electrical transport of AB-BA domain walls. We investigated the temperature dependence of electrical transport both through AB-stacked bilayer graphene and through AB-BA domain walls. Extended Data Fig. 3 displays representative data for CNP resistance at $V_{bg} = -140$ V for a bilayer graphene device without a domain wall (black line) and a device with a domain wall (red line). For the bilayer graphene bulk device without a domain wall, the CNP resistance increases by over 300 kΩ (2.5 times) as the temperature is varied from 50 K to 1.8 K. This large increase of resistance at lower temperatures is characteristic of insulating behaviour. In comparison, the CNP resistance of device with a domain wall (red line) has much weaker temperature dependence: the resistance increase is only 5 kΩ (~50%), and part of the increase is due to the contribution from bulk bilayer graphene channel in parallel with the one-dimensional domain wall channel. The weak temperature dependence of conductance along the domain wall is characteristic of a metallic system.

31. Lopes dos Santos, J. M. B., Peres, N. M. R. & Castro Neto, A. H. Graphene bilayer with a twist: electronic structure. *Phys. Rev. Lett.* **99,** 256802 (2007).
32. Mak, K. F., Sfeir, M. Y., Misewich, J. A. & Heinz, T. F. The evolution of electronic structure in few-layer graphene revealed by optical spectroscopy. *Proc. Natl Acad. Sci. USA* **107,** 14999–15004 (2010).
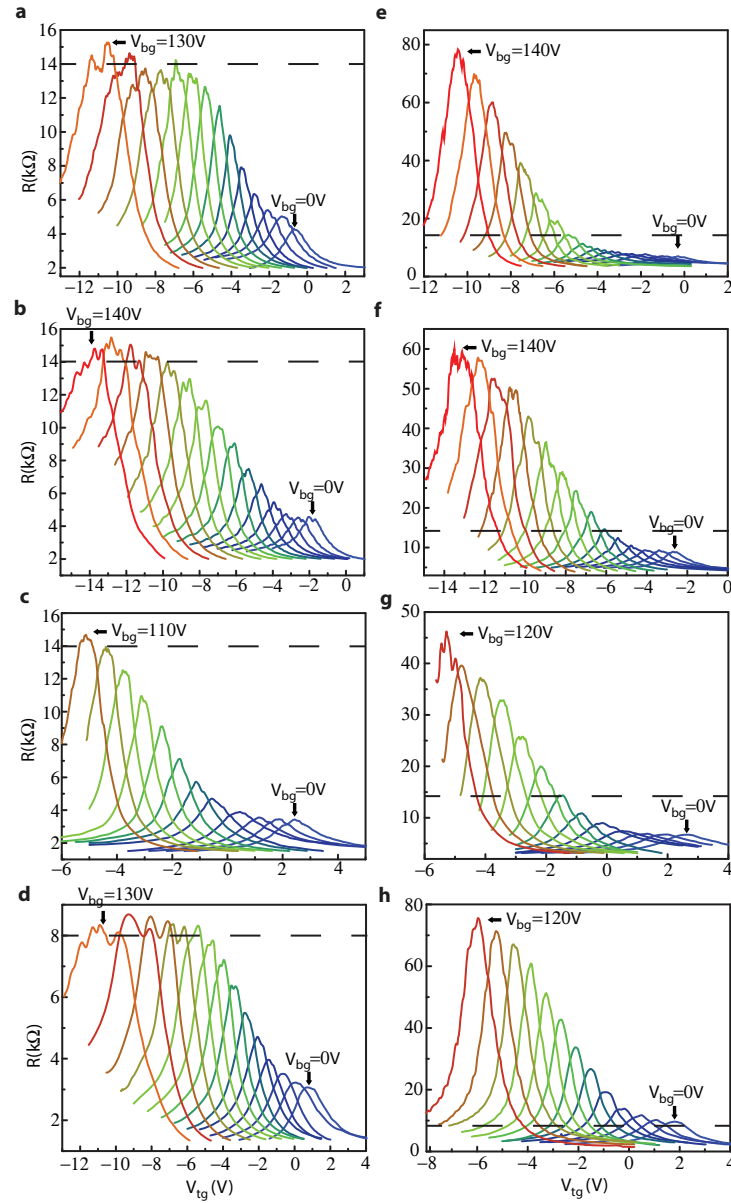
Extended Data Figure 1 | Electrical transport of dual-gated bilayer graphene devices with (left) and without (right) AB-BA domain walls. **a–d**, Gate-dependent resistance R for three bilayer graphene devices with one domain wall (**a–c**) and one with two domain walls (**d**) at 4.2 K. The backgate voltage V_{bg} is varied from 0 V to 140 V with a step of 10 V, and the top gate voltage V_{tg} is swept continuously. The resistance peak in each trace corresponds to a CNP. The resistance at the CNP first increases and then saturates at

$\sim$14 kΩ in **a–c** ($\sim$8 kΩ in **d**). The dashed lines indicate the saturated conductance values. **e–h**, Gate-dependent resistance of four reference bilayer graphene devices without domain walls. The resistance at the CNP increases continuously at high V_{bg} owing to the opening of a bandgap, and reaches values much higher than that of corresponding devices with bilayer graphene domain walls (indicated by the dashed line).
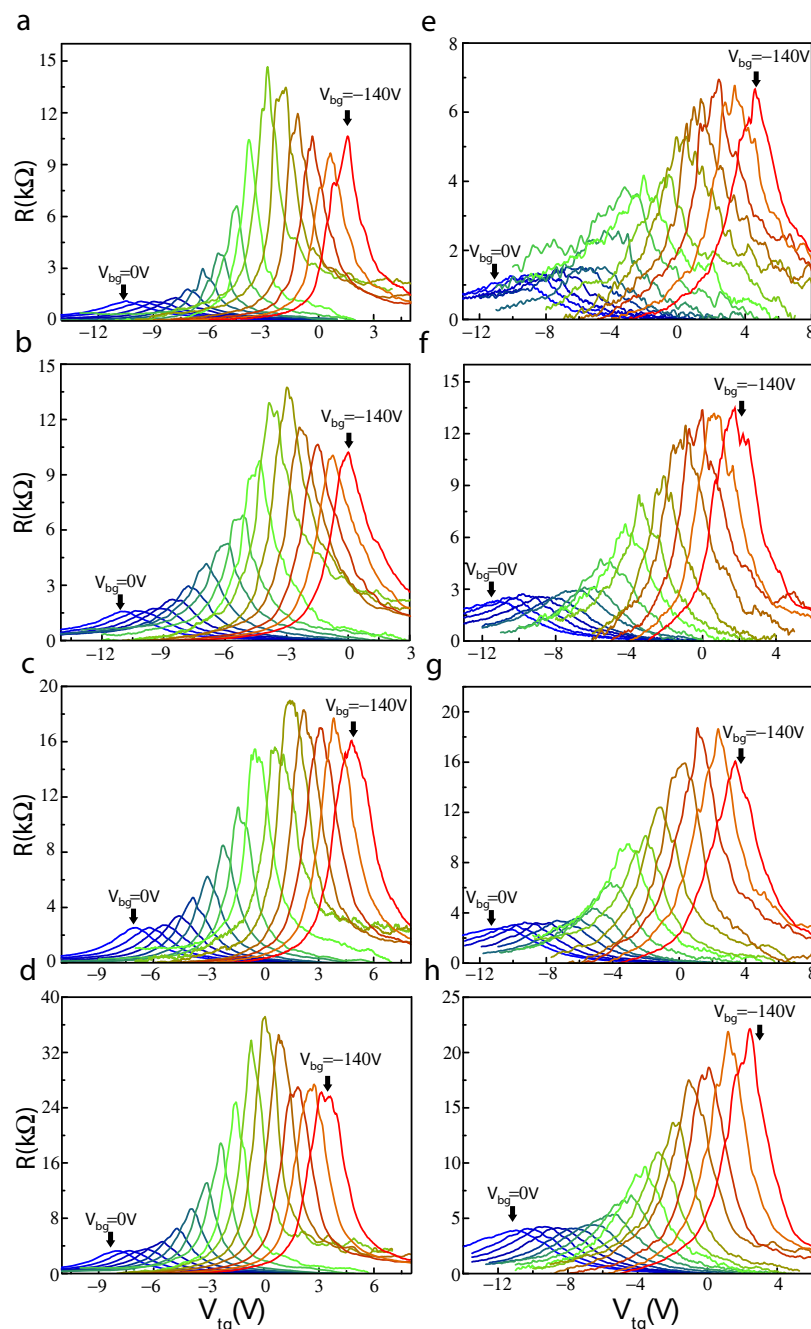
Extended Data Figure 2 | Length dependence of electrical transport of AB-BA domain walls. Data in left panels are from one graphene flake, data in right panels are from another. **a–d**, Gate-dependent resistance for four bilayer graphene devices fabricated on a long domain wall with channel lengths of 200 nm, 400 nm, 600 nm and 800 nm (defined by the top gate width), respectively. **e–h**, Gate-dependent resistance for four bilayer graphene devices fabricated on another long domain wall with channel lengths of 200 nm,

400 nm, 600 nm and 800 nm respectively. The backgate voltage V_{bg} is varied from 0 to -140 V with a step of 10 V, and the top gate voltage V_{tg} is swept continuously. The resistance peak in each trace corresponds to a CNP. The resistance at the CNP first increases as the bandgap is opened in the bulk and then saturates at high V_{bg}, where the electrical transport is dominated by one-dimensional channels along the domain wall.
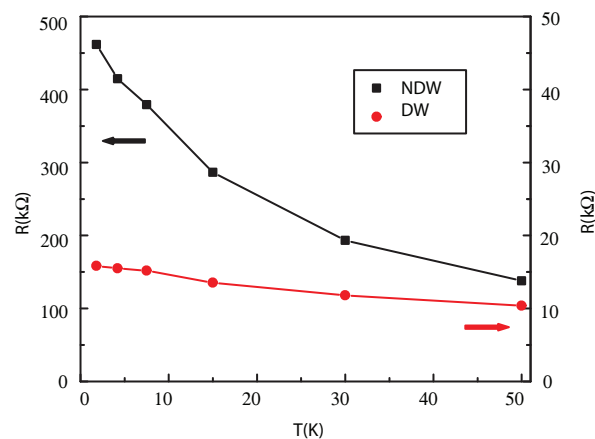
Extended Data Figure 3 | Temperature-dependent electrical transport through bulk bilayer graphene and through the AB-BA domain wall. The CNP resistance ($V_{bg} = -140$ V) of a bilayer graphene device without a domain wall (NDW, black line) increases by over 300 kΩ (~2.5 times) as the temperature is decreased from 50 K to 1.8 K. In contrast, the CNP resistance ($V_{bg} = -140$ V) of a bilayer graphene device with domain wall (red line, the same device as in Extended Data Fig. 2c) increases by only 5 kΩ (~50%) at the same temperature. The much weaker temperature dependence of the domain wall resistance arises from a metallic one-dimensional conducting channel.

High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity

Kibum Kang[1]*, Saien Xie[2]*, Lujie Huang[1], Yimo Han[2], Pinshane Y. Huang[2], Kin Fai Mak[3,4], Cheol-Joo Kim[1], David Muller[2,3] & Jiwoong Park[1,3]

The large-scale growth of semiconducting thin films forms the basis of modern electronics and optoelectronics. A decrease in film thickness to the ultimate limit of the atomic, sub-nanometre length scale, a difficult limit for traditional semiconductors (such as Si and GaAs), would bring wide benefits for applications in ultrathin and flexible electronics, photovoltaics and display technology[1–3]. For this, transition-metal dichalcogenides (TMDs), which can form stable three-atom-thick monolayers[4], provide ideal semiconducting materials with high electrical carrier mobility[5–10], and their large-scale growth on insulating substrates would enable the batch fabrication of atomically thin high-performance transistors and photodetectors on a technologically relevant scale without film transfer. In addition, their unique electronic band structures provide novel ways of enhancing the functionalities of such devices, including the large excitonic effect[11], bandgap modulation[12], indirect-to-direct bandgap transition[13], piezoelectricity[14] and valleytronics[15]. However, the large-scale growth of monolayer TMD films with spatial homogeneity and high electrical performance remains an unsolved challenge. Here we report the preparation of high-mobility 4-inch wafer-scale films of monolayer molybdenum disulphide (MoS$_2$) and tungsten disulphide, grown directly on insulating SiO$_2$ substrates, with excellent spatial homogeneity over the entire films. They are grown with a newly developed, metal–organic chemical vapour deposition technique, and show high electrical performance, including an electron mobility of $30\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ at room temperature and $114\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ at 90 K for MoS$_2$, with little dependence on position or channel length. With the use of these films we successfully demonstrate the wafer-scale batch fabrication of high-performance monolayer MoS$_2$ field-effect transistors with a 99% device yield and the multi-level fabrication of vertically stacked transistor devices for three-dimensional circuitry. Our work is a step towards the realization of atomically thin integrated circuitry.

Existing growth methods for large-scale monolayer TMDs have so far produced materials with limited spatial uniformity and electrical performance. For instance, the sulphurization of metal or metal compounds only provides control over the average layer number, producing spatially inhomogeneous mixtures of monolayer, multi-layer and no-growth regions[16,17]. Although chemical vapour deposition (CVD) based on solid-phase precursors (such as MoO$_3$, MoCl$_5$ or WO$_3$)[18–23] has shown better thickness control on a large scale, the electrical performance of the resulting material, which is often reported from a small number of devices in selected areas, fails to show spatially uniform high carrier mobility.

Here we report the growth of semiconducting monolayer films of MoS$_2$ and tungsten disulphide (WS$_2$) on silicon oxide on a 4-inch wafer scale, with both excellent electrical performance and structural continuity, maintained uniformly over the entire films. Figure 1 presents our continuous TMD monolayer films and shows their wafer-scale homogeneity and intrinsic optical properties. The colour photos of MoS$_2$ (Fig. 1a; greenish yellow) and WS$_2$ (Fig. 1b; yellow) films

grown on a transparent 4-inch fused silica wafer show that the TMD grown region (right half) is uniform over the whole substrate and clearly distinguishable from the bare silica substrate (left half). The optical absorption, photoluminescence and Raman spectra measured from our films show characteristics unique to monolayer MoS$_2$ and WS$_2$, respectively (Fig. 1d–f). All of these measured spectra have the same peak positions as in exfoliated monolayer samples (denoted by diamonds)[13,24–26], regardless of the location of the measurements within our films (Supplementary Fig. 1). The X-ray photoelectron spectra taken from our monolayer MoS$_2$ film show almost identical features to those of bulk single crystal with a low level of defects, further confirming the precise chemical composition and the high quality of our MoS$_2$ film (Supplementary Fig. 2).

Figure 1c shows a photo of a MoS$_2$ film grown on a 4-inch SiO$_2$/Si wafer. The monolayer film was patterned using standard photolithography and oxygen plasma etching to form MoS$_2$-covered squares (dark, 6 mm wide) with an array of 3 μm holes. An enlarged, normalized optical reflection image (Fig. 1h) displays a homogeneous reflection contrast for the entire MoS$_2$-covered region, confirming uniform monolayer growth everywhere with no gaps. In addition, Fig. 1g shows a scanning electron microscope (SEM) image of an array of fully suspended monolayer MoS$_2$ membranes (2 μm in diameter) fabricated by transferring our metal–organic chemical vapour deposition (MOCVD)-grown film onto a SiN grid with holes. Its high fabrication yield (>99.5%) suggests mechanical strength and continuity of the film. The widefield photoluminescence images of these films (insets to Fig. 1g, h) show strong, spatially uniform photoluminescence signals, further confirming that they are continuous monolayer MoS$_2$, with its high quality maintained even after patterning or transfer. The same spatial uniformity was seen in the optical reflection and photoluminescence images of a monolayer WS$_2$ film that was similarly grown and patterned (Fig. 1i). Together, the data in Fig. 1 confirm that our MoS$_2$ and WS$_2$ films are continuous monolayers, spatially uniform over the entire 4-inch growth substrates with intrinsic optical properties. Below, using MoS$_2$ as the main example, we discuss the growth (Fig. 2) and the excellent electrical properties (Fig. 3) of these MOCVD-grown films.

Figure 2a schematically explains our MOCVD growth, where we only use gas-phase precursors of Mo(CO)$_6$, W(CO)$_6$, (C$_2$H$_5$)$_2$S and H$_2$, all diluted in argon as a carrier gas (see Supplementary Methods). The concentration of each reactant can be precisely controlled during the entire growth time by regulating the partial pressure (P_X) of each reactant X. Thus our setup offers an ideal environment for maximizing the areal coverage of the monolayer and for engineering the film structure by controlling the nucleation density and intergrain stitching. Figure 2 summarizes our key findings.

First, our MoS$_2$ film is grown in the layer-by-layer growth mode, which is ideal for uniform layer control over the large scale. Figure 2c plots the areal coverage of monolayer (θ_{1L}) and multilayer ($\theta_{\geq 2L}$; mostly bilayer) regions measured from our MoS$_2$ grown on SiO$_2$/Si;

[1]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. [2]School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA. [3]Kavli Institute at Cornell for Nanoscale Science, Ithaca, New York 14853, USA. [4]Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA.
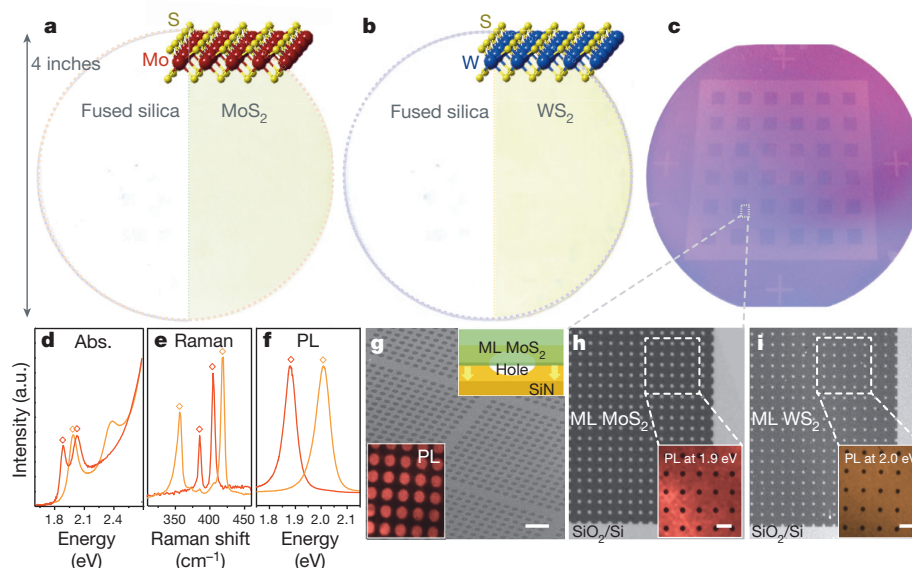*These authors contributed equally to this work.

Figure 1 | Wafer-scale monolayer TMD films. a, b, Photographs of monolayer MoS_2 (**a**) and WS_2 (**b**) films grown on 4-inch fused silica substrates, with diagrams of their respective atomic structures. The left halves show the bare fused silica substrate for comparison. **c,** Photograph of patterned monolayer MoS_2 film on a 4-inch SiO_2/Si wafer (the darker areas are covered by MoS_2). **d,** Optical absorption spectra of MOCVD-grown monolayer MoS_2 (red line) and WS_2 (orange line) films in the photon energy range from 1.6 to 2.7 eV. **e,** Raman spectra of as-grown monolayer MoS_2 and WS_2, normalized to the silicon peak intensity. **f,** Normalized photoluminescence spectra of as-grown monolayer MoS_2 and WS_2. The peak positions in **d–f** are consistent with those seen from exfoliated samples (diamonds). **g,** SEM image and photoluminescence (PL) image (bottom inset, at 1.9 eV) of monolayer (ML) MoS_2 membranes suspended over a SiN TEM grid with 2 μm holes (a diagram of the suspended film is shown in the top inset). Scale bar, 10 μm. **h, i,** Optical images (normalized to the bare substrate region) of patterned monolayer MoS_2 (**h**) and WS_2 (**i**) on SiO_2, taken from the wafer-scale patterned films. The insets show photoluminescence images at 1.9 eV (**h**) and 2.0 eV (**i**). Scale bars, 10 μm.

Fig. 2b shows optical images at different growth times, revealing the initial nucleation on the SiO_2 surface ($t = 0.5t_0$), subsequent monolayer growth near ($0.8t_0$) and at the maximum monolayer coverage (t_0), followed by nucleation mainly at grain boundaries ($1.2t_0$) and bilayer growth ($2t_0$). We observed no nucleation of a second layer while the first layer was forming ($\theta_{\geq 2L} \approx 0$ when $t < t_0$), producing an optimal growth time t_0 near full monolayer coverage ($\theta_{1L} \approx 1$). Additional photoluminescence and electron microscope images taken after different growth times further suggested that the edge attachment was the main mechanism for the monolayer growth after nucleation and that the neighbouring monolayer grains were uniformly connected by tilt grain boundaries with enhanced photoluminescence[19] at $t = t_0$ (see Supplementary Notes and Supplementary Figs 3 and 4). The standard thin-film growth model[27] suggests that this growth mode is effective below a certain deposition rate of the growth species, above which it suggests a different mode that forms thicker islands. Indeed, the layer-by-layer growth of MoS_2 film was observed only when we applied a low partial pressure ($P_{Mo} \approx 10^{-4}$ Torr in Fig. 2b, c) of Mo vapour (produced by the thermal decomposition of $Mo(CO)_6$; see Supplementary Notes and Supplementary Fig. 5) in the presence of excess $(C_2H_5)_2S$. In contrast, the growth at a higher P_{Mo} was no longer in the layer-by-layer growth mode, instead simultaneously producing a mixture of monolayer, multilayer and no-growth regions (Supplementary Fig. 6). For the uniform monolayer growth over a large substrate, it is thus important to maintain a low P_{Mo} constantly over the entire growth region and over time, the key technical capability provided by our MOCVD setup (see Supplementary Fig. 7 for the spatially homogeneous monolayer nucleation on a multi-inch scale).

Second, the grain structure of our MoS_2 film, including the average grain size and the intergrain connection, depends sensitively on the concentrations of H_2, $(C_2H_5)_2S$ as well as residual water. As a representative example, Fig. 2d shows the two main effects of H_2, whose presence is necessary for removing carbonaceous species generated during the MOCVD growth: the average grain size increases from hundreds of nanometres to more than 10 μm with decreasing H_2 flow, and the MoS_2 grains grown under higher H_2 flow (Fig. 2d, right image)

have mostly perfect triangular shapes without merging with neighbouring grains, a trend that disappears with lower H_2 flow (left and middle images). These observations are consistent with the H_2-induced decomposition of $(C_2H_5)_2S$ (increasing nucleation due to hydrogenolysis)[28] and the etching of the MoS_2 (preventing intergrain connection)[29] as reported previously. (For further discussion on the effects of $(C_2H_5)_2S$ and water, see Supplementary Notes and Supplementary Fig. 6.) To grow continuous monolayer MoS_2 with a large grain size and high-quality intergrain stitching, we thus flowed optimal amounts of H_2 and $(C_2H_5)_2S$ and dehydrated the growth environment.

The darkfield transmission electron microscope (DF-TEM) and annular darkfield scanning TEM (ADF-STEM) images shown in Fig. 2e, f confirm the structural continuity of our MoS_2 film grown under those conditions on the nanometre and atomic length scales. The DF-TEM image shows a continuous polycrystalline monolayer film with no visible gaps and a bilayer area of less than 0.5%. Further analysis of the DF-TEM and electron diffraction data (see Supplementary Fig. 8) confirms a uniform angular distribution of crystal orientations with no preferred intergrain tilt angle for grain boundaries. The ADF-STEM data (Fig. 2f; more images are shown in Supplementary Fig. 9) further confirm that adjacent grains are likely to be connected by a high-quality lateral connection with structures similar to those seen in previous reports[18,19]. The MoS_2 films shown in Fig. 1 as well as those whose electrical properties we show in Figs 3 and 4 were grown under the conditions described in Supplementary Methods, producing an average grain size of ~1 μm (see Fig. 2b, e). Almost identical growth parameters with $P_W \approx 10^{-4}$ Torr produced monolayer WS_2 films as shown in Fig. 1b, i, indicating the same layer-by-layer growth for WS_2 with a similar t_0.

The electrical properties of our monolayer MoS_2 films have two important characteristics: the spatial uniformity over a large scale and excellent transport properties similar to those seen in exfoliated samples. All our electrical measurements in Figs 3 and 4 (except those in Fig. 3c) were performed at room temperature. Figure 3a shows a plot of sheet conductance ($\sigma_\square$) against backgate voltage (V_{BG}) measured from a monolayer MoS_2 field-effect transistor (FET; optical image
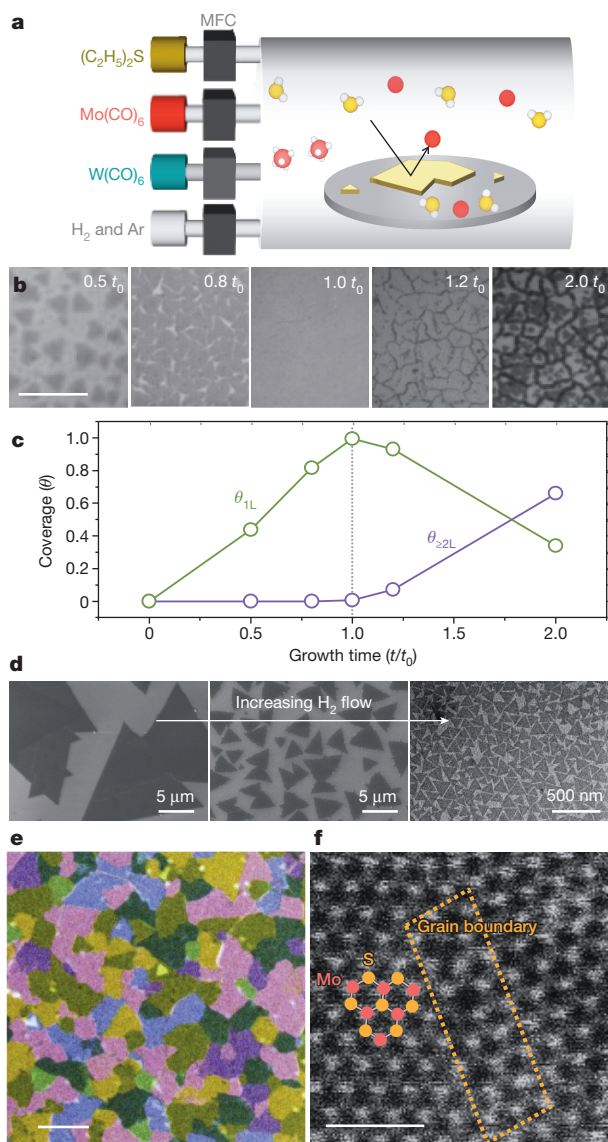
Figure 2 | MOCVD growth of continuous monolayer MoS$_2$ film. a, Diagram of our MOCVD growth setup. Precursors were introduced to the growth setup with individual mass flow controllers (MFCs). Red, Mo or W atom; yellow, S atom; white, carbonyl or ethyl ligand. **b**, Optical images of MOCVD-grown MoS$_2$ at the indicated growth times, where t_0 was the optimal growth time for full monolayer coverage. Scale bar, 10 μm. **c**, Coverage ratio for monolayer (θ_{1L}, green) and multi-layer ($\theta_{\geq 2L}$, purple) regions as a function of growth time. **d**, Grain size variation of monolayer MoS$_2$ depending on the hydrogen flow rate; from left to right, 5 standard cm^3 min^{-1} (sccm) (SEM image shown), 20 sccm (SEM) and 200 sccm (TEM). **e**, False-colour DF-TEM image showing a continuous monolayer MoS$_2$ film. Scale bar, 1 μm. **f**, ADF-STEM image of a laterally stitched grain boundary in a monolayer MoS$_2$ film, with red and yellow dots representing the Mo and S atoms, respectively. Scale bar, 1 nm.

shown in the inset) with multiple electrodes for the four-probe measurements (except for a channel length L of 34 μm). It includes several curves for different L ranging between 1.6 and 34 μm (shifted from the bottom curve for clarity), all of which show nearly identical behaviours, including the n-type conductance, carrier concentration ($\sim 4 \times 10^{12}$ cm^{-2} at $V_{BG} = 0$ V) and high field-effect mobility (μ_{FE}). Figure 3b further plots μ_{FE} measured from five such devices, fabricated at random locations and separated by up to 3.3 mm on a single chip. All the devices show similar μ_{FE} near 30 cm^2 V^{-1} s^{-1}, independent of L and device location, with similarly uniform $\sigma_\square - V_{BG}$ curves (shown in Supplementary Fig. 10), suggesting the spatial homogeneity of the

electrical properties of the MoS$_2$ film at length scales ranging from micrometres to millimetres.

The distribution of μ_{FE} of our devices is compared with the results of multiple devices from two previous reports, each measured from individual grains of exfoliated[7] or CVD-grown[21] MoS$_2$ samples. We find that μ_{FE} measured from our MOCVD film is similar to the median μ_{FE} (denoted by a star) of exfoliated samples (and several times higher than the CVD results) and has a much narrower distribution. In addition, the temperature dependence of μ_{FE} (Fig. 3c) measured from the same device in Fig. 3a shows higher μ_{FE} at lower temperatures (92 cm^2 V^{-1} s^{-1} at 100 K) and intrinsic, phonon-limited electron transport similar to the behaviours previously observed in exfoliated samples (data from ref. 6 shown in Fig. 3c) but different from those observed from a CVD sample with stronger effects from defects[30]. Specifically, our data show that the temperature dependence of mobility follows a power law of $\mu_{FE} \approx T^{-\gamma}$ with exponent $\gamma = 1.6$ for temperatures between 150 and 300 K, close to the value (1.69) predicted by theory[5] and consistent with results from previous experiments (average value ranging between 0.6 and 1.7)[6-9] for a similar temperature range. Finally, Fig. 3d shows a high-performance MoS$_2$ FET fabricated with an individual top-gate electrode (V_{TG}). It has a high on/off conductance ratio ($\sim 10^6$), current saturation at relatively low bias V_{SD} (lower inset to Fig. 3d), high field-effect mobility (~ 29 cm^2 V^{-1} s^{-1}) and large transconductance (~ 2 μS μm^{-1}), all of which are comparable to the best reported results[6-8]. We note that our devices studied in Fig. 3a–d were fabricated at random locations using a polycrystalline monolayer MoS$_2$ film, unlike the devices with single-grain samples used for comparison. In addition, the electrical properties measured from a separate monolayer MoS$_2$ film with a larger average grain size of 3 μm (instead of 1 μm in Fig. 3) have almost identical characteristics, including the channel-length independence of μ_{FE} and the phonon-limited transport at $T > 150$ K (see Supplementary Fig. 11; with the low-temperature mobility as high as 114 cm^2 V^{-1} s^{-1} at 90 K). Taken together, our data confirm the spatial uniformity and high electrical performance of our MoS$_2$ FETs independent of the average grain size, which suggests that the intergrain boundaries in our film do not significantly degrade their electrical transport properties. This is probably due to the formation of well-stitched intergrain boundaries with a low level of defects, an explanation also supported by the ADF-STEM (Fig. 2f) and X-ray photoelectron spectrum data (Supplementary Fig. 2) discussed above. Our data therefore lead us to conclude that our optimized MOCVD growth provides an electrically homogeneous monolayer MoS$_2$ film. Moreover, we successfully fabricated and measured 60 FETs by using a monolayer WS$_2$ film. Even though the growth of monolayer WS$_2$ was not carefully optimized, these devices showed excellent electrical properties, with their μ_{FE} as high as 18 cm^2 V^{-1} s^{-1} at room temperature (Fig. 3e) and a median μ_{FE} close to 5 cm^2 V^{-1} s^{-1}. In addition, the WS$_2$ device in Fig. 3e showed a high on/off ratio of 10^6 and the current saturation behaviour (inset to Fig. 3e) as in our MoS$_2$ devices. (See Supplementary Fig. 12 for data from additional monolayer WS$_2$ FET devices).

The structural and electrical uniformity of our MoS$_2$ film enables the wafer-scale batch fabrication of high performance FETs as demonstrated in Fig. 3f, g. Figure 3f shows a photo of 8,100 MoS$_2$ FETs with a global back gate, which were fabricated on a 4-inch SiO$_2$/Si wafer with a standard photolithography process. The middle and bottom insets to Figure 3f show colour-scale maps of $\sigma_\square$ measured from 100 MoS$_2$ FETs in one square region at $V_{BG} = +50$ V and -50 V, respectively; the top inset to Fig. 3f shows an enlarged optical image of the devices. We observed an almost perfect device yield of 99%; only two out of 200 FETs that we characterized (including data from an adjacent region) did not conduct. Our data also confirm the spatially uniform n-type transistor operation (larger $\sigma_\square$ for positive V_{BG}) with similar V_{BG} dependence for all our devices and high on-state device conductance. We further observed similarly uniform V_{BG} dependence from FET devices fabricated using monolayer MoS$_2$ films with different average grain sizes, as characterized by the histograms of the threshold voltages
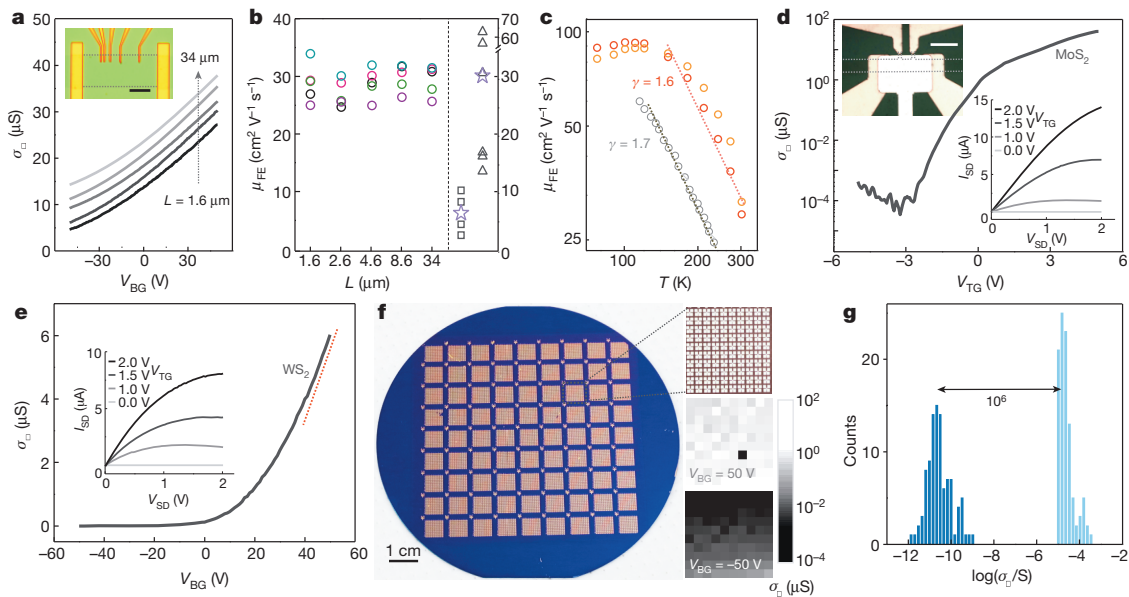
Figure 3 | Electrical characterization and batch fabrication of monolayer TMD FETs. a, Gate-dependent sheet conductance ($\sigma_\square$) of monolayer MoS$_2$ FETs measured at different channel lengths L (curves displaced from the bottom for clarity). Inset: optical image of the device; scale bar, 10 μm. **b**, Field-effect mobility (μ_{FE}) measured from five MoS$_2$ FETs (ellipses) fabricated at random locations with different L. Data from previous results for CVD-grown samples[21] (squares) and exfoliated samples[7] (triangles) are shown for comparison (purple stars indicate their medians). **c**, Temperature dependence of μ_{FE} measured from the device in Fig. 3a at $L = 1.6$ μm (orange) and 8.6 μm (red), and from a previous report on exfoliated samples[6] (grey), both showing phonon-limited intrinsic transport. **d**, Top gate (V_{TG})-dependent $\sigma_\square$ for dual-gate monolayer MoS$_2$ FET (device shown in the upper inset). Lower inset: V_{TG}-dependent I_{SD}–V_{SD} curves showing current saturation and ohmic electrode contact. Scale bar, 10 μm. **e**, Gate-dependent $\sigma_\square$ of a monolayer WS$_2$ FET showing $\mu_{FE} = 18$ cm^2 V^{-1} s^{-1}. Inset: V_{TG}-dependent I_{SD}–V_{SD} curves showing current saturation and ohmic electrode contact. **f**, Batch-fabricated 8,100 MoS$_2$ FET devices on a 4-inch SiO$_2$/Si wafer. Top inset: enlarged image of one square containing 100 devices. Middle and bottom insets: corresponding colour maps of $\sigma_\square$ at gate bias $V_{BG} = 50$ V and -50 V, respectively, with the black block in the middle inset representing the only non-conducting device. **g**, Histogram of on-state and off-state $\sigma_\square$ of 100 dual-gate FETs showing a median on–off ratio of 10^6 and a high on-state conductivity. Dark blue, $V_{TG} = -5$ V; pale blue, $V_{TG} = +5$ V. All measurements were performed at room temperatures except those in **c**.

(Supplementary Fig. 13). Similarly, we fabricated 100 individually addressable dual-gate MoS$_2$ FETs (similar to the device in Fig. 3d) on another wafer piece. The histogram of the on-state $\sigma_\square$ ($V_{TG} = 5$ V; median carrier concentration $\sim 7 \times 10^{12}$ cm^{-2}) and off-state $\sigma_\square$ ($V_{TG} = -5$ V) collected from all such FETs (Fig. 3g) shows strong peaks above 10^{-5} S and near 10^{-11} S, respectively, confirming a uniform conductance switching behaviour with high on-state $\sigma_\square$ (>10 μS) and on–off ratio ($\sim 10^6$). In addition, most of these batch-fabricated FETs had a high μ_{FE} (>10 cm^2 V^{-1} s^{-1}; see Supplementary Fig. 14).

The data presented in Figs 1–3 confirm the structural and electrical uniformity of the wafer-scale monolayer MoS$_2$ film grown by our MOCVD method. This makes our film compatible with batch device fabrication processes on a technologically relevant scale. Moreover, because SiO$_2$ provides a substrate for its growth, one can produce high-quality monolayer films on a variety of substrates by depositing SiO$_2$ before the growth. This versatility would permit the fabrication of high-performance FETs directly on non-conventional substrates, such as metal and thermally stable plastic. In addition, one can integrate multiple layers of MoS$_2$ devices by repeating the TMD film growth, device fabrication and SiO$_2$ deposition, which could enable novel three-dimensional circuitry.

In Fig. 4 we demonstrate this unique potential by producing multi-stacked monolayer MoS$_2$ films as well as electronic devices fabricated at different vertical levels. Figure 4a shows diagrams and photographs of three substrates each with single, double or triple monolayer MoS$_2$ films grown at different levels. The first (bottom) monolayer film was grown on a fused silica substrate; the additional layers were grown on SiO$_2$ (100 nm thick) deposited on the previously grown MoS$_2$ monolayer by using plasma-enhanced CVD. The colour of the substrate, which remains uniform for each substrate, becomes darker as the number of layers increases. Their absorption spectra, shown in Fig. 4b, present almost identical absorption at all

measured wavelengths, once normalized by the number of stacks grown (see the inset), suggesting little degradation of the optical properties of the monolayer MoS$_2$ films after subsequent oxide deposition and MoS$_2$ growth.

Figure 4c shows successive diagrams of our multi-stacked device fabrication process: growth of the first MoS$_2$ monolayer on a SiO$_2$/Si wafer, FET fabrication, deposition of SiO$_2$ (thickness 500 nm; see Supplementary Methods for details), and growth of the second MoS$_2$ monolayer and FET fabrication. A false-colour SEM image in Fig. 4d shows an array of MoS$_2$ FETs successfully fabricated with this process. It includes functioning MoS$_2$ FETs located at two different vertical levels whose conductance can be simultaneously modulated with a global back gate. The I_{SD}–V_{SD} curves measured from two FETs, adjacent both laterally and vertically (see the inset to Fig. 4d), are shown in Fig. 4e. Both devices show a V_{BG}-dependent conductance change (notice the smaller change for the second layer) with an on-state $\sigma_\square$ of 2.5 μS (first layer) and 1.5 μS (second layer), respectively. Furthermore, we measured similar μ_{FE} values (11.5 and 8.8 cm^2 V^{-1} s^{-1}) from the two devices (Supplementary Fig. 15). The two monolayer MoS$_2$ films were grown on SiO$_2$ substrates prepared differently, and the first-layer device had gone through additional steps, including the second MoS$_2$ growth. Our data in Fig. 4 thus confirm the compatibility of our MOCVD-grown MoS$_2$ films with conventional thin-film deposition and multi-stacking, which could be used to develop a three-dimensional device architecture based on TMD.

Our high-mobility monolayer TMD films can be used immediately for the batch fabrication of TMD-based integrated circuitry consisting of FETs, photodetectors and light emitting diodes, on a technologically relevant multi-inch wafer scale. In addition, because our MOCVD growth is controlled by the kinetics of precursor supply rather than specific precursor–substrate chemistry
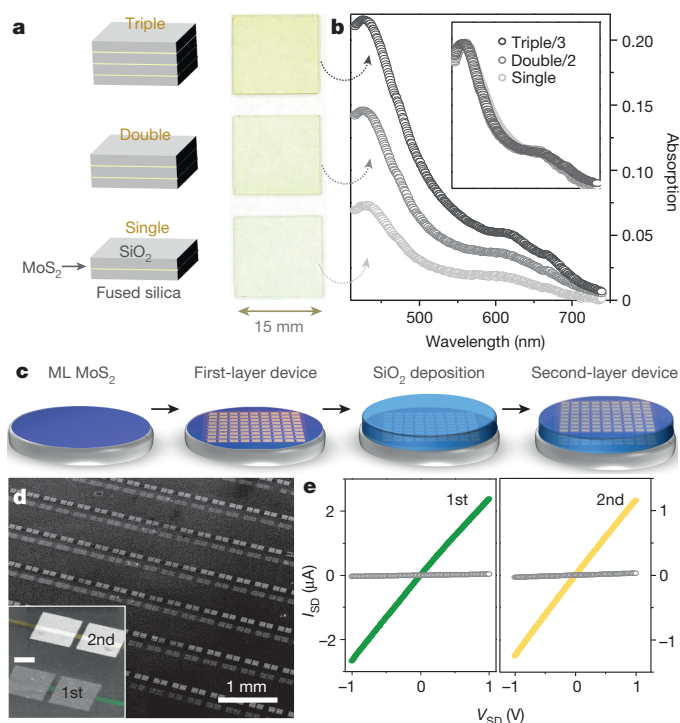
Figure 4 | Multi-stacking of MoS₂/SiO₂ structure. a, Diagrams (left) and optical images (right) of single, double and triple stacking of monolayer MoS₂/SiO₂. **b**, Optical absorption spectra for single, double and triple stacks, respectively (normalized spectra shown in the inset). **c**, Diagram for fabrication of MoS₂ device/SiO₂ stacking by alternating MOCVD growth, device fabrication with photolithography, and SiO₂ deposition. See the text for details. **d**, False-colour SEM image of MoS₂ FET arrays on first (bottom) and second (top) layers (the inset shows an enlarged image of a pair of devices in the same relative positions as in the main panel; scale bar, 50 μm). **e**, I_{SD}–V_{SD} curves measured from two neighbouring devices on the first (left) and second (right) layers, both showing n-type conductance switching. For the first layer, $V_{BG} = 50$ V (green) and -50 V (grey); for the second layer, $V_{BG} = 100$ V (yellow) and -100 V (grey).

(an example of the latter would be the different graphene growth modes on copper and nickel), its use is not limited to the TMD–substrate combinations reported here. Instead, it could be generalized for producing various TMD materials, both semiconductor (for example MoSe₂ or WTe₂) and metal (for example NbSe₂ or TaS₂), with precise layer control over a large scale. Indeed, our data show that, as an initial demonstration, monolayer TMD growth is possible on a variety of other technologically important substrates (such as Al₂O₃, SiN and HfO₂) with the same growth conditions as those developed for SiO₂ (see Supplementary Fig. 16 for MoS₂ growth and device fabrication on these substrates, using these non-optimized conditions). Our versatile MOCVD growth therefore provides a new avenue for the growth, patterning and integration of multiple high-quality monolayer TMD films with different compositions and electrical properties on a single substrate, enabling the future development of atomically thin integrated circuitry.

1. Rogers, J. A., Someya, T. & Huang, Y. Materials and mechanics for stretchable electronics. *Science* **327**, 1603–1607 (2010).
2. Ko, H. *et al.* Ultrathin compound semiconductor on insulator layers for high-performance nanoscale transistors. *Nature* **468**, 286–289 (2010).
3. Mitzi, D. B., Kosbar, L. L., Murray, C. E., Copel, M. & Afzali, A. High-mobility ultrathin semiconducting films prepared by spin coating. *Nature* **428**, 299–303 (2003).
4. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499**, 419–425 (2013).
5. Kaasbjerg, K., Thygesen, K. S. & Jacobsen, K. W. Phonon-limited mobility in n-type single-layer MoS₂ from first principles. *Phys. Rev. B* **85**, 115317 (2012).
6. Baugher, B. W. H., Churchill, H. O. H., Yang, Y. & Jarillo-Herrero, P. Intrinsic electronic transport properties of high-quality monolayer and bilayer MoS₂. *Nano Lett.* **13**, 4212–4216 (2013).
7. Radisavljevic, B. & Kis, A. Mobility engineering and a metal–insulator transition in monolayer MoS₂. *Nature Mater.* **12**, 815–820 (2013).
8. Schmidt, H. *et al.* Transport properties of monolayer MoS₂ grown by chemical vapor deposition. *Nano Lett.* **14**, 1909–1913 (2014).
9. Jariwala, D. *et al.* Band-like transport in high mobility unencapsulated single-layer MoS₂ transistors. *Appl. Phys. Lett.* **102**, 173107 (2013).
10. Ovchinnikov, D. *et al.* Electrical transport properties of single-layer WS₂. *ACS Nano* **8**, 8174–8181 (2014).
11. Ye, Z. *et al.* Probing excitonic dark states in single-layer tungsten disulphide. *Nature* **513**, 214–218 (2014).
12. Feng, Q. *et al.* Growth of large-area 2D MoS₂₍₁₋ₓ₎Se₂ₓ semiconductor alloys. *Adv. Mater.* **26**, 2648–2653 (2014).
13. Mak, K. F., Lee, C., Hone, J., Shan, J. & Heinz, T. F. Atomically thin MoS₂: a new direct-gap semiconductor. *Phys. Rev. Lett.* **105**, 136805 (2010).
14. Wu, W. *et al.* Piezoelectricity of single-atomic-layer MoS₂ for energy conversion and piezotronics. *Nature* **514**, 470–474 (2014).
15. Mak, K. F., McGill, K. L., Park, J. & McEuen, P. L. The valley Hall effect in MoS₂ transistors. *Science* **344**, 1489–1492 (2014).
16. Lin, Y.-C. *et al.* Wafer-scale MoS₂ thin layers prepared by MoO₃ sulfurization. *Nanoscale* **4**, 6637–6641 (2012).
17. Song, I. *et al.* Patternable large-scale molybdenium disulfide atomic layers grown by gold-assisted chemical vapor deposition. *Angew. Chem. Int. Ed.* **53**, 1266–1269 (2014).
18. Najmaei, S. *et al.* Vapour phase growth and grain boundary structure of molybdenum disulphide atomic layers. *Nature Mater.* **12**, 754–759 (2013).
19. Van der Zande, A. M. *et al.* Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide. *Nature Mater.* **12**, 554–561 (2013).
20. Lee, Y.-H. *et al.* Synthesis and transfer of single-layer transition metal disulfides on diverse surfaces. *Nano Lett.* **13**, 1852–1857 (2013).
21. Liu, H. *et al.* Statistical study of deep submicron dual-gated field-effect transistors on monolayer chemical vapor deposition molybdenum disulfide films. *Nano Lett.* **13**, 2640–2646 (2013).
22. Yu, Y. *et al.* Controlled scalable synthesis of uniform, high-quality monolayer and few-layer MoS₂ films. *Sci. Rep.* **3**, 1866 (2013).
23. Zhang, Y. *et al.* Controlled growth of high-quality monolayer WS₂ layers on sapphire. *ACS Nano* **7**, 8963–8971 (2013).
24. Zhao, W. *et al.* Evolution of electronic structure in atomically thin sheets of WS₂ and WSe₂. *ACS Nano* **7**, 791–797 (2013).
25. Li, H. *et al.* From bulk to monolayer MoS₂: evolution of Raman scattering. *Adv. Funct. Mater.* **22**, 1385–1390 (2012).
26. Berkdemir, A. *et al.* Identification of individual and few layers of WS₂ using Raman spectroscopy. *Sci. Rep.* **3**, 1755 (2013).
27. King, D. A. & Woodruff, D. P. *Growth and Properties of Ultrathin Epitaxial Layers* (Elsevier, 1997).
28. Connor, R. & Adkins, H. Hydrogenolysis of oxygenated organic compounds. *J. Am. Chem. Soc.* **54**, 4678–4690 (1932).
29. Cristol, S. *et al.* Theoretical study of the MoS₂ (100) surface: a chemical potential analysis of sulfur and hydrogen coverage. *J. Phys. Chem. B* **104**, 11220–11229 (2000).
30. Najmaei, S. *et al.* Electrical transport properties of polycrystalline monolayer molybdenum disulfide. *ACS Nano* **8**, 7930–7937 (2014).

Author Contributions K.K. and S.X. contributed equally to this work. K.K, S.X. and J.P. conceived the experiments. K.K. and S.X. performed the synthesis, optical characterization, device fabrication, and electrical measurements. L.H. conducted low-temperature electrical measurement, with assistance from K.F.M, Y.H, P.Y.H. and D.A.M. performed atomic resolution STEM imaging. K.K. carried out DF-TEM and data analysis, with assistance from C.-J.K. K.K, S.X. and J.P. wrote the manuscript. All authors discussed the results and commented on the manuscript.

LETTER

Precise interpolar phasing of abrupt climate change during the last ice age

WAIS Divide Project Members*

The last glacial period exhibited abrupt Dansgaard–Oeschger climatic oscillations, evidence of which is preserved in a variety of Northern Hemisphere palaeoclimate archives[1]. Ice cores show that Antarctica cooled during the warm phases of the Greenland Dansgaard–Oeschger cycle and vice versa[2,3], suggesting an interhemispheric redistribution of heat through a mechanism called the bipolar seesaw[4–6]. Variations in the Atlantic meridional overturning circulation (AMOC) strength are thought to have been important, but much uncertainty remains regarding the dynamics and trigger of these abrupt events[7–9]. Key information is contained in the relative phasing of hemispheric climate variations, yet the large, poorly constrained difference between gas age and ice age and the relatively low resolution of methane records from Antarctic ice cores have so far precluded methane-based synchronization at the required sub-centennial precision[2,3,10]. Here we use a recently drilled high-accumulation Antarctic ice core to show that, on average, abrupt Greenland warming leads the corresponding Antarctic cooling onset by 218 ± 92 years (2σ) for Dansgaard–Oeschger events, including the Bølling event; Greenland cooling leads the corresponding onset of Antarctic warming by 208 ± 96 years. Our results demonstrate a north-to-south directionality of the abrupt climatic signal, which is propagated to the Southern Hemisphere high latitudes by oceanic rather than atmospheric processes. The similar interpolar phasing of warming and cooling transitions suggests that the transfer time of the climatic signal is independent of the AMOC background state. Our findings confirm a central role for ocean circulation in the bipolar seesaw and provide clear criteria for assessing hypotheses and model simulations of Dansgaard–Oeschger dynamics.

Net heat transport by the Atlantic branch of the global ocean overturning circulation is northwards at all latitudes, resulting in a heat flux from the Southern Hemisphere (SH) to the Northern Hemisphere (NH)[4]. Variations in this flux act to redistribute heat between the hemispheres, a mechanism commonly invoked to explain abrupt sub-orbital climatic variability and the asynchronous coupling of Greenland and Antarctic temperature variations on these timescales[2,6]. Millennial-scale AMOC variability is corroborated by North Atlantic proxies for deep-water ventilation and provenance that suggest decreased North Atlantic deep water production and the intrusion of southern-sourced water masses during stadial (that is, North Atlantic cold) periods[11,12]. The oceanic instabilities are accompanied by shifting atmospheric transport patterns. Proxy data and climate models consistently indicate northward and southward migrations of the intertropical convergence zone in response to abrupt NH warming and cooling, respectively[13,14]. Abrupt NH events may also induce changes in the strength and position of the SH mid-latitude westerlies. Strengthened and/or southward-shifted westerlies during NH stadials have the potential to warm the Southern Ocean and Antarctica by enhancing the wind-driven upwelling of relatively warm circumpolar deep waters, providing a direct atmospheric pathway for the bipolar seesaw to operate[15]. Climate model simulations further suggest that atmospheric readjustment in response to decreased North Atlantic

deep water formation can induce inhomogeneous temperature changes over the Antarctic continent[16,17], possibly by means of wind-driven changes in sea-ice distribution[8].

Atmospheric teleconnections operate on seasonal to decadal timescales because of the fast response time of the atmosphere[8,18,19]. By contrast, oceanic teleconnections can operate on a wide range of timescales from decadal to multi-millennial, depending on the processes and ocean basins involved[19–21]. Climatic signals can be rapidly propagated from the North Atlantic to the South Atlantic via Kelvin waves[5], but models suggest a more gradual (centennial timescale) propagation from the South Atlantic to the SH high latitudes as a result of the absence of a zonal topographic boundary at the latitudes of the Antarctic circumpolar current (ACC)[19,21]. The timescale on which newly formed North Atlantic deep water is exported to the Southern Ocean is similarly around several centuries[22]. The bipolar seesaw relationship observed between ice core records potentially bears the imprint of both atmospheric and oceanic teleconnections; precise constraints on the interhemispheric phasing can help distinguish which mode dominates, and can also identify leads and lags[10,19,21].

Completed in December 2011, the West Antarctic Ice Sheet (WAIS) Divide ice core (WDC)[17] was drilled and recovered to a depth of 3,405 m. Here we present results from the deep part of the core (Fig. 1), extending back to 68 kyr before 1950 (before present; BP). The WDC climatic records show no stratigraphic disturbances (such as large-scale folds) that are commonly encountered near the base of ice cores, a fact we attribute to basal melting at the site that removes old ice before such disturbances can develop and that decreases shear stress by lubricating the bed. The Greenland NGRIP core similarly has strong basal melting and an undisturbed stratigraphy[1]. WDC δ18O of ice, a proxy for local condensation temperature, exhibits clear millennial-scale variability as observed across the Antarctic continent (Fig. 1c, d). We evaluate the phasing of WDC millennial variability relative to δ18O of the Greenland NGRIP core (Fig. 1a). For each Greenland Dansgaard–Oeschger (DO) event we can identify a corresponding Antarctic Isotopic Maximum (AIM) event[3], although AIM 9 is only very weakly expressed in WDC (we adopt the naming convention whereby AIM x is concomitant with DO x).

To investigate the phasing of the bipolar seesaw, we have synchronized WDC to the Greenland NGRIP core by means of globally well-mixed atmospheric methane (CH_4; Fig. 1b)[2,23]. WDC allows CH_4 synchronization at unprecedented, sub-centennial precision as a result of the small difference between gas age and ice age (Δage) and continuous, centimetre-scale resolution CH_4 record (Extended Data Figs 1 and 2). Because uncertainties in the relative phasing of CH_4 and Greenland climate[24] are smaller than uncertainties in Greenland Δage, we synchronize WDC CH_4 directly to NGRIP δ18O (rather than to NGRIP CH_4)[23].

As pointed out by several authors[6,21], a strong anti-correlation exists between Greenland δ18O and the rate of change (that is, the first time derivative) of Antarctic δ18O. This relationship is consistent with a simple thermodynamic model of the bipolar seesaw in which Antarctic temperature variations are moderated by the large thermal mass of the
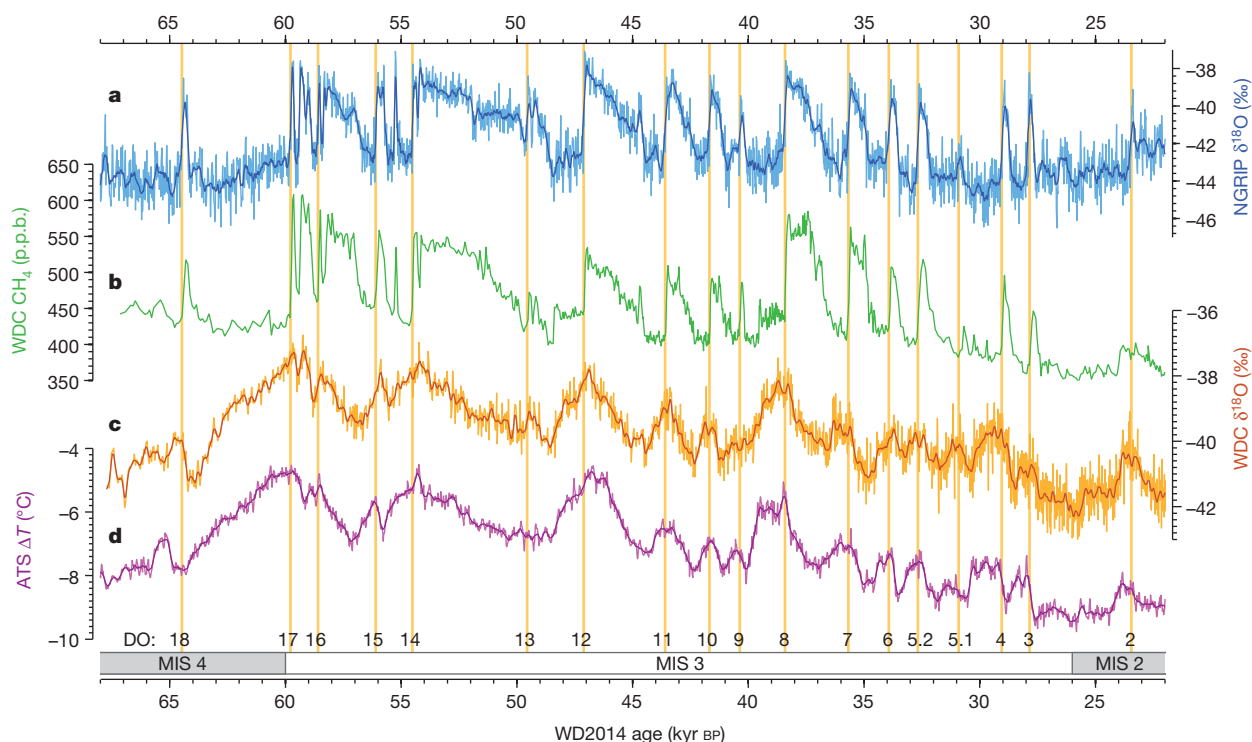
Figure 1 | Records of glacial abrupt millennial-scale climatic variability.
a, Greenland NGRIP $\delta^{18}O$ record[1] on GICC05 $\times$ 1.0063 chronology
(Methods). **b**, WDC discrete CH_4 record on the WD2014 chronology, which is
based on layer counting (0–31.2 kyr) and CH_4 synchronization to NGRIP
(31.2–68 kyr)[23]. **c**, WDC $\delta^{18}O$ record. **d**, Antarctic temperature stack (ATS)[30] in
degrees Celsius relative to the present day on AICC12 $\times$ 1.0063 chronology.
DO/AIM events are indicated with orange vertical bars, numbered at the
bottom of the figure.

Southern Ocean[6]. In evaluating this relationship at WAIS Divide
(Fig. 2a, blue curve), we find the strongest anti-correlation at a NH
lead of about 170 years (blue dot in Fig. 2a). This timing is distinct from
the zero-year lead commonly assumed in mathematical descriptions of
the bipolar seesaw. The analysis is repeated using an ensemble of
4×10^3 alternative WDC chronologies (Methods), showing that the
centennial-scale timing is a robust result (Fig. 2d) with an estimated
uncertainty of 69 years (2σ uncertainty bounds are used throughout
this work). Evaluation of the cross-correlations using WDC CH_4
instead of NGRIP $\delta^{18}O$ gives almost identical results (Fig. 2a, green
curve), in line with the notion that CH_4 is a good proxy for Greenland
climate[2].

Next we analyse the DO–AIM coupling in more detail. The bipolar
seesaw theory suggests that during Greenland stadial periods heat
accumulates in the SH, causing gradual warming of Antarctica; during
interstadial periods Antarctica is cooling. The abrupt stadial–intersta-
dial transitions in Greenland are accompanied by a breakpoint in the
WDC $\delta^{18}O$ record, where the warming trend changes to a cooling
trend. Although the timing of abrupt transitions can be established
unambiguously in both CH_4 and Greenland $\delta^{18}O$ records, the smaller
signal-to-noise ratio of Antarctic $\delta^{18}O$ time series complicates break-
point identification for individual AIM events. To overcome this limi-
tation, we stack and average the AIM events to detect their shared
climatic signal (Methods and Extended Data Figs 3 and 4). We align
WDC records at the midpoint of the DO CH_4 transitions, which is set
to lag the midpoint in the NGRIP $\delta^{18}O$ transition by 56 ± 38 years[24].
The relative timing of the WDC $\delta^{18}O$ and CH_4 curves is determined by
Δage, which is the largest source of uncertainty (Methods).

We find that Antarctic cooling is delayed relative to abrupt NH
warming by 218 ± 92 years on average (Fig. 2b, e); similarly,
Antarctic warming lags NH cooling by 208 ± 96 years (Fig. 2c, f).
The robustness of this result is demonstrated with a Monte Carlo
analysis in which we randomly perturb the relative alignment of the
individual AIM events, in combination with an ensemble of 4×10^3

alternative WDC chronologies. The timing of the WDC $\delta^{18}O$ break-
point (orange dots in Fig. 2b, c) is determined by using an automated
fitting algorithm (Methods and Extended Data Fig. 5). Performing the
analysis separately on a stack of only the major AIM events (AIM 4, 8,
12, 14 and 17), only the minor AIM events or stacks of eight randomly
selected AIM events gives nearly identical phasing (Extended Data Fig.
6), excluding the possibility that the result is dominated by the timing
of a few prominent events. We repeat the stacking procedure for the
WDC sea-salt sodium record (Extended Data Fig. 7), and find that this
tracer, which has been interpreted as a proxy for sea-ice production[25],
changes almost synchronously with WDC $\delta^{18}O$—possibly reflecting a
common forcing by Southern Ocean temperatures[6].

At the onset of the NH Bølling warm period (DO 1; 14.6 kyr) we find
a north–south phasing comparable to that of the glacial period, with a
256 ± 133-year lag of the Antarctic $\delta^{18}O$ response (Fig. 3 and
Extended Data Table 1). The interpolar phasing at the onset and
termination of the Younger Dryas stadial is ambiguous in our record
(Extended Data Table 1), mostly because of the relatively gradual
nature of the Younger Dryas onset and the presence of two local
maxima in the WAIS-D $\delta^{18}O$ record around the Younger Dryas–
Holocene transition (possibly reflecting regional climate). A recent
study using a CH_4-synchronized stack of five near-coastal Antarctic
cores spanning the deglaciation suggested a near synchrony of
Antarctic breakpoints and Greenland transitions within a dating
uncertainty of 200 years[10]. Our result overlaps with this range.
However, the much smaller dating uncertainty at WDC and our use
of the full sequence of glacial AIM events shows that a centennial-scale
Antarctic lag is a systematic feature of DO–AIM coupling throughout
the glacial period.

The lead of NH climate revealed at WDC (Fig. 2) provides clues
about DO climate dynamics. The simplest interpretation is that the
abrupt phases of the DO cycle are initiated in the NH, presumably in
the North Atlantic. This is consistent with several proposed DO
mechanisms, such as North Atlantic sea-ice dynamics[9], freshwater
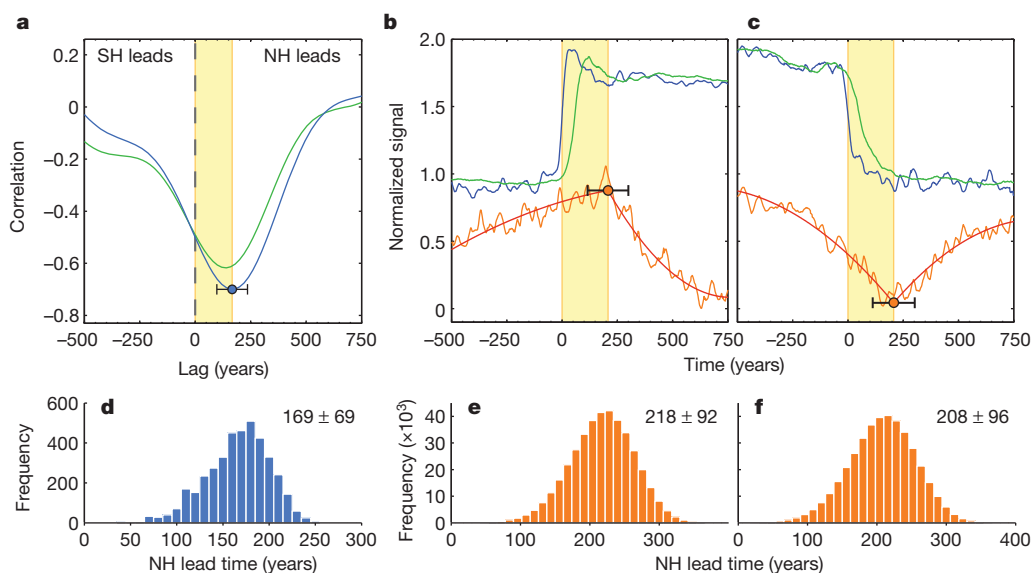
Figure 2 | Interhemispheric phasing of the bipolar seesaw. a, Lagged correlation between NGRIP $\delta^{18}O$ and WDC $d(\delta^{18}O)/dt$ (blue), and between WDC CH_4 and $d(\delta^{18}O)/dt$ (green). Millennial-scale variability is isolated by using a fourth-order Butterworth bandpass filter (500–10,000-year window); the CH_4-synchronized part of the records is used (31.2–68 kyr). **b**, DO 3–18 stack of NGRIP $\delta^{18}O$ (blue), WDC CH_4 (green) and WDC $\delta^{18}O$ (orange with curve fit), aligned at the midpoint of the DO warming signal. Events are averaged with their original amplitudes and normalized after stacking for convenience of visualization. **c**, As in **b**, but for NH abrupt cooling events (that is, the interstadial terminations). **d–f**, Histograms of NH lead time associated with **a–c**, respectively, generated by binning solutions from the sensitivity study. The total number of solutions is 4×10^3 in **d**, and 4×10^5 in **e** and **f**. Distribution mean and 2σ probability bounds are listed in the panels. Shaded vertical yellow bars (upper panels) show NH lead time; the error bar represents 2σ as defined for the lower panels.

forcing[8] or ice shelf collapse[7]. We acknowledge, however, that the abrupt North Atlantic events could be the response to a remote process not visible in the ice core records; therefore, although we cannot ascertain the location of the elusive DO 'trigger' (if such a concept is even meaningful in a highly coupled dynamical system), our results clearly indicate a north-to-south directionality of the abrupt phases of the bipolar seesaw signal.

The centennial timescale of the NH lead demonstrates the dominance of oceanic processes in propagating NH temperature anomalies to the SH high latitudes. Any atmospheric teleconnection would be manifested within at most a few decades after the abrupt NH event[18]; on this timescale the WDC $\delta^{18}O$ stacks (Fig. 2b, c) show no discernible temperature response above the noise. We estimate the noise level by subtracting the fitted curves from the $\delta^{18}O$ stacks; the remaining signal has a 2σ variability of 0.14‰, or about 0.18 °C assuming an isotope sensitivity of $0.8‰ K^{-1}$. We therefore estimate an upper bound of 0.18 °C on an atmospherically induced Antarctic temperature response (for comparison, millennial Antarctic temperature variations are on the order of 1–2 °C). A readjustment of atmospheric transport may induce spatially inhomogeneous temperature changes[8,16] that may be (partly) responsible for the heterogeneity in expression, or shape, of AIM events across Antarctica[3].

We find that on average the DO cooling signal is transmitted as fast to Antarctica as the DO warming signal is (our sensitivity study suggests a difference in propagation time of 10 ± 89 years). This implies that the north-to-south propagation time is independent of the AMOC background state; that is, it is independent of whether the AMOC is in the weak or strong overturning state. Modelling work suggests that the meridional propagation time across the ACC latitudes depends primarily on ACC strength[21]; our inference of similar propagation times may thus reflect a stability of the ACC on millennial timescales. There is also a conspicuous synchronicity between the phasing of the bipolar seesaw and the duration of the abrupt increase in CO_2 at 14.6 kyr (Fig. 3d); if the former does indeed reflect the timescale of the oceanic response, this may hold clues about the (unidentified) source of this CO_2 (ref. 26).

Proxy records of North Atlantic ventilation and overturning strength during Marine Isotope Stage 3 typically show the most prominent excursions during Heinrich stadials[11,12], periods of extreme cold in the North Atlantic associated with layers of ice-rafted debris in ocean sediments that represent times of massive delivery of
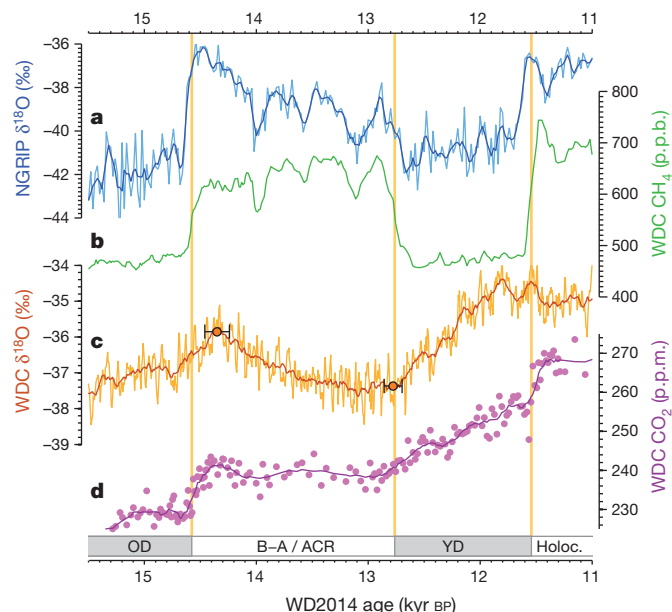
Figure 3 | Timing of the last deglaciation. a, NGRIP $\delta^{18}O$ on GICC05 chronology[1]. **b**, WDC CH_4. **c**, WDC $\delta^{18}O$ with breakpoints as orange dots and error bars showing the 2σ uncertainty bounds (Extended Data Table 1). **d**, WDC CO_2 data (dots) with spline fit (solid line)[26]. Period abbreviations: OD, Oldest Dryas; B–A, Bølling–Allerød; ACR, Antarctic Cold Reversal; YD, Younger Dryas; Holoc., Holocene. Vertical orange lines correspond to the midpoints of the WDC CH_4 transitions. NGRIP and WDC chronologies are both based on annual-layer counting, and are fully independent.

debris-laden icebergs (Heinrich events). This suggests that there may be a difference between climate dynamics during Heinrich and non-Heinrich stadials. We find no anomalous NH lead time for DO transitions directly after or before Heinrich stadials (Extended Data Fig. 6), and thus no evidence that Heinrich stadials are unusual from the perspective of the oceanic teleconnections that dominate the bipolar seesaw. Similarly, previous work has shown that an increase in Antarctic temperature is linearly related to Greenland stadial duration, irrespective of the occurrence of Heinrich events during these stadials[3]. Recent studies suggest that decreases in AMOC strength precede the Heinrich events, allowing the possibility that the latter are a response to AMOC variations rather than their cause[27,28]. The main influence of Heinrich events on the bipolar seesaw may thus be to lengthen the stadial periods during which they occur by suppressing the AMOC by means of iceberg-delivered freshwater, allowing large amounts of heat to build up in the Southern Ocean.

Although both the data[5] and the models[19,21] suggest a fast temperature response in the South Atlantic (decadal-scale in models), there is currently no consensus in existing model studies on the physical mechanisms and timescales of the processes that propagate temperature signals between the South Atlantic and the Southern Ocean; complicating factors to consider include the lack of a zonal boundary to support the propagation of Kelvin waves[21], the large thermal inertia of the Southern Ocean[6], the complexity of eddy heat transport across the Antarctic circumpolar current, and the coupling between the extent of Antarctic sea ice and the overturning circulation[29]. In this context, our precise phasing observations provide a new constraint with which to test future model simulations seeking to capture the dynamics of millennial-scale climate variability.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. NGRIP Project Members. High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431,** 147–151 (2004).
2. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291,** 109–112 (2001).
3. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444,** 195–198 (2006).
4. Crowley, T. J. North Atlantic Deep Water cools the southern hemisphere. *Paleoceanography* **7,** 489–497 (1992).
5. Barker, S. *et al.* Interhemispheric Atlantic seesaw response during the last deglaciation. *Nature* **457,** 1097–1102 (2009).
6. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18,** 1087 (2003).
7. Petersen, S. V., Schrag, D. P. & Clark, P. U. A new mechanism for Dansgaard–Oeschger cycles. *Paleoceanography* **28,** 24–30 (2013).
8. Rind, D. *et al.* Effects of glacial meltwater in the GISS coupled atmosphere–ocean model. 2. A bipolar seesaw in Atlantic Deep Water production. *J. Geophys. Res.* **106** (D21) 27355–27365 (2001).
9. Dokken, T. M., Nisancioglu, K. H., Li, C., Battisti, D. S. & Kissel, C. Dansgaard–Oeschger cycles: interactions between ocean and sea ice intrinsic to the Nordic seas. *Paleoceanography* **28,** 491–502 (2013).
10. Pedro, J. B. *et al.* The last deglaciation: timing the bipolar seesaw. *Clim. Past.* **7,** 671–683 (2011).
11. Piotrowski, A. M., Goldstein, S. L., Hemming, S. R., Fairbanks, R. G. & Zylberberg, D. R. Oscillating glacial northern and southern deep water formation from combined neodymium and carbon isotopes. *Earth Planet. Sci. Lett.* **272,** 394–405 (2008).
12. Martrat, B. *et al.* Four climate cycles of recurring deep and surface water destabilizations on the Iberian margin. *Science* **317,** 502–507 (2007).
13. Peterson, L. C., Haug, G. H., Hughen, K. A. & Röhl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the last glacial. *Science* **290,** 1947–1951 (2000).
14. Chiang, J. C. H. & Bitz, C. M. Influence of high latitude ice cover on the marine Intertropical Convergence Zone. *Clim. Dyn.* **25,** 477–496 (2005).
15. Toggweiler, J. R. & Lea, D. W. Temperature differences between the hemispheres and ice age climate variability. *Paleoceanography* **25,** PA2212 (2010).
16. Timmermann, A. *et al.* Towards a quantitative understanding of millennial-scale Antarctic warming events. *Quat. Sci. Rev.* **29,** 74–85 (2010).
17. WAIS Divide Project Members. Onset of deglacial warming in West Antarctica driven by local orbital forcing. *Nature* **500,** 440–444 (2013).
18. Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. *Clim. Change* **54,** 251–267 (2002).
19. Liu, Z. Y. & Alexander, M. Atmospheric bridge, oceanic tunnel, and global climatic teleconnections. *Rev. Geophys.* **45,** RG2005 (2007).
20. Masuda, S. *et al.* Simulated rapid warming of abyssal North Pacific waters. *Science* **329,** 319–322 (2010).
21. Schmittner, A., Saenko, O. A. & Weaver, A. J. Coupling of the hemispheres in observations and simulations of glacial climate change. *Quat. Sci. Rev.* **22,** 659–671 (2003).
22. England, M. H. The age of water and ventilation timescales in a global ocean model. *J. Phys. Oceanogr.* **25,** 2756–2777 (1995).
23. Buizert, C. *et al.* The WAIS Divide deep ice core WD2014 chronology. Part 1. Methane synchronization (68–31 ka BP) and the gas age–ice age difference. *Clim. Past.* **11,** 153–173 (2015).
24. Baumgartner, M. *et al.* NGRIP CH4 concentration from 120 to 10 kyr before present and its relation to δ^{15}N temperature reconstruction from the same ice core. *Clim. Past* **10,** 903–920 (2014).
25. Wolff, E. W., Rankin, A. M. & Röthlisberger, R. An ice core indicator of Antarctic sea ice production? *Geophys. Res. Lett.* **30,** 2158 (2003).
26. Marcott, S. A. *et al.* Centennial-scale changes in the global carbon cycle during the last deglaciation. *Nature* **514,** 616–619 (2014).
27. Gutjahr, M. & Lippold, J. Early arrival of Southern Source Water in the deep North Atlantic prior to Heinrich event 2. *Paleoceanography* **26,** PA2101 (2011).
28. Marcott, S. A. *et al.* Ice-shelf collapse from subsurface warming as a trigger for Heinrich events. *Proc. Natl Acad. Sci. USA* **108,** 13415–13419 (2011).
29. Ferrari, R. *et al.* Antarctic sea ice control on ocean circulation in present and glacial climates. *Proc. Natl Acad. Sci. USA* **111,** 8753–8758 (2014).
30. Parrenin, F. *et al.* Synchronous change of atmospheric CO2 and Antarctic temperature during the last deglacial warming. *Science* **339,** 1060–1063 (2013).

WAIS Divide Project Members

Christo Buizert[1], Betty Adrian[2], Jinho Ahn[3], Mary Albert[4], Richard B. Alley[5], Daniel Baggenstos[6], Thomas K. Bauska[1], Ryan C. Bay[7], Brian B. Bencivengo[2], Charles R. Bentley[8], Edward J. Brook[1], Nathan J. Chellman[9], Gary D. Clow[10], Jihong Cole-Dai[11], Howard Conway[12], Eric Cravens[13], Kurt M. Cuffey[14], Nelia W. Dunbar[15], Jon S. Edwards[1], John M. Fegyveresi[5], Dave G. Ferris[11], Joan J. Fitzpatrick[16], T. J. Fudge[12], Chris J. Gibson[8], Vasileios Gkinis[17,18], Joshua J. Goetz[8], Stephanie Gregory[4], Geoffrey M. Hargreaves[2], Nels Iverson[15], Jay A. Johnson[8], Tyler R. Jones[17], Michael L. Kalk[1], Matthew J. Kippenhan[19], Bess G. Koffman[20], Karl Kreutz[21], Tanner W. Kuhl[8], Donald A. Lebar[8], James E. Lee[1], Shaun A. Marcott[1,22], Bradley R. Markle[12], Olivia J. Maselli[9], Joseph R. McConnell[9], Kenneth C. McGwire[9], Logan E. Mitchell[1], Nicolai B. Mortensen[8], Peter D. Neff[23], Kunihiko Nishiizumi[24], Richard M. Nunn[2], Anais J. Orsi[6,25], Daniel R. Pasteris[9], Joel B. Pedro[18,26], Erin C. Pettit[27], P. Buford Price[7], John C. Priscu[28], Rachael H. Rhodes[1], Julia L. Rosen[1], Andrew J. Schauer[12], Spruce W. Schoenemann[12], Paul J. Sendelbach[8], Jeffrey P. Severinghaus[6], Alexander J. Shturmakov[8], Michael Sigl[9], Kristina R. Slawny[8], Joseph M. Souney[29], Todd A. Sowers[5], Matthew K. Spencer[30], Eric J. Steig[12], Kendrick C. Taylor[9], Mark S. Twickler[29], Bruce H. Vaughn[17], Donald E. Voigt[5], Edwin D. Waddington[12], Kees C. Welten[24], Anthony W. Wendricks[8], James W. C. White[17], Mai Winstrup[12,18], Gifford J. Wong[31] & Thomas E. Woodruff[32]

[1]College of Earth, Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. [2]US Geological Survey National Ice Core Laboratory, Denver, Colorado 80225, USA. [3]School of Earth and Environmental Science, Seoul National University, Seoul 151-742, Korea. [4]Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755, USA. [5]Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [6]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093, USA. [7]Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. [8]Ice Drilling Design and Operations, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [9]Desert Research Institute, Nevada System of Higher Education, Reno, Nevada 89512, USA. [10]US Geological Survey, Boulder, Colorado 80309, USA. [11]Department of Chemistry and Biochemistry, South Dakota State University, Brookings, South Dakota 57007, USA. [12]Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195-1310, USA. [13]ADC Management Services, Lakewood, Colorado 80226, USA. [14]Department of Geography, University of California at Berkeley, Berkeley, California 94709, USA. [15]Earth and Environmental Science Department, New Mexico Tech, Socorro, New Mexico 87801, USA. [16]US Geological Survey, Denver, Colorado 80225, USA. [17]Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80309-0450, USA. [18]Centre for ice and climate, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark. [19]Antarctic Support Contract, Lockheed Martin US Antarctic Program, Centennial, Colorado 80112, USA. [20]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA. [21]Climate Change Institute and School of Earth and Climate Sciences, University of Maine, Orono, Maine 04469, USA. [22]University of Wisconsin-Madison, Madison, Wisconsin, Wisconsin 53706 USA. [23]Antarctic Research Centre, Victoria University of Wellington, Wellington 6012, New Zealand. [24]Space Sciences Laboratory, University of California at Berkeley, Berkeley, California 94720, USA. [25]Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre Simon Laplace, 91191 Gif-Sur-Yvette, France. [26]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington 98195, USA. [27]Department of Geosciences, University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA. [28]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana 59717, USA. [29]Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA. [30]School of Physical Sciences, Lake Superior State University, Sault Sainte Marie, Michigan 49783, USA. [31]Department of Earth Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA. [32]PRIME Laboratory, Purdue University, West Lafayette, Indiana 47907, USA.

METHODS

Core recovery and processing. The location of the WAIS Divide ice core (79.48° S, 112.11° W) was selected to provide the highest possible time-resolution record of Antarctic climate during the past 50 kyr or more, and to ensure that the record from the last deglaciation would not be in ice from the lower-quality brittle-ice zone[17]. The site has a present-day ice accumulation rate of 22 cm yr^{-1}, divide flow and an ice thickness of 3,460 m (J. Paden, personal communication 2011). No stratigraphic irregularities were detected in the core using visual observation, multitrack electrical conductivity measurements or gas stratigraphy. Coring was stopped 50 m above the bed to prevent contamination of the unfrozen basal environment. Drilling of the main core (WDC06A) started during the 2006/2007 field season and finished during the 2011/2012 field season, at the final depth of 3,405 m. Drilling was done with the US Deep Ice Sheet Coring (DISC) drill, a new drill designed and operated by the Ice Drilling Design and Operations group (engineering team at University of Wisconsin, Madison). Core handling and distribution of ice samples were performed by the National Ice Core Laboratory (US Geological Survey) and the Science Coordination Office (University of New Hampshire and Desert Research Institute). The drill and core handling methods had many innovations that improved core quality, especially in the brittle ice. These included improved mechanical characteristics of the drill, better control of the drilling operation, core handling procedures that minimized thermal and mechanical shock to core, netting to contain the brittle ice, and temperature tracking of all ice samples from the drill to the laboratories.

Data description. Water $^{18}O/^{16}O$ composition ($\delta^{18}O$) was measured at IsoLab, University of Washington, Seattle, Washington. Measurement procedures for WDC have been described elsewhere[17,31]. In short, measurements were made at 0.5 m depth averaged resolution, using laser spectroscopy (Picarro L2120-i water isotope analyser). $\delta^{18}O$ data are reported relative to the VSMOW (Vienna Standard Mean Ocean water) standard, and normalized to the VSMOW-SLAP (Standard Light Antarctic Precipitation) scale.

Two separate data sets of atmospheric methane exist for WDC. The first is based on discrete sampling of the core; it was measured jointly at Penn State University (0–68 kyr, 0.5–2 m resolution) and at Oregon State University (11.4–24.8 kyr, 1–2 m resolution). For this record the air was extracted using a melt–refreeze technique from ~50 g ice samples, and analysed on a gas chromatograph with a flame ionization detector[32]. Corrections for gas solubility, blank and gravitational enrichment for the Oregon State University data were performed in accordance with ref. 33; corrections to the Penn State University data are described in ref. 17. A second data set (R. H. Rhodes, E. J. Brook, J. C. H. Chiang, T. Blunier, O. J. Maselli, J. R. McConnell, D. Romanini and J. P. Severinghaus, unpublished observations) is based on continuous flow analysis (CFA) in combination with laser spectroscopy[34,35]. This data set, which is of higher temporal resolution, was used to determine the timing of the abrupt DO transitions (Extended Data Fig. 2). All CH$_4$ data are reported on the NOAA04 scale[36].

WDC Na concentrations were measured at the Ultra Trace Chemistry Laboratory at the Desert Research Institute, Reno, Nevada, by means of CFA. Sticks of ice were melted continuously on a heated plate. Potentially contaminated meltwater from the outer part of the ice stick was discarded. Two inductively coupled plasma mass spectrometers were used to analyse the stream of melt-water[37]. The CFA chemistry and CH$_4$ data were measured simultaneously on the same ice samples, and all measurements are co-registered in depth.

Age scale and Δage. For WDC we use the WD2014 chronology, which is based on annual-layer counting down to 2,850 m (31.2 kyr BP), and on stratigraphic matching by means of globally well-mixed methane for the deepest part of the core (2,850–3,405 m)[23]. The deep section of the WD2014 chronology has been synchronized to a linearly scaled version of the layer-counted Greenland ice core chronology (GICC05). Multiplying the GICC05 chronology by 1.0063 brings the ages of abrupt DO events as observed in Greenland in agreement with the ages of these events observed in the U/Th-dated Hulu Cave speleothem record (which provides better absolute age control)[23]. The NGRIP data in Fig. 1a have similarly been placed on the 1.0063 × GICC05 age scale; NGRIP data in Fig. 3a are plotted on the original GICC05 chronology.

WDC permits precise interhemispheric synchronization for two reasons. First, as a result of the high accumulation rates, Δage remains relatively small. WDC Δage is about an order of magnitude smaller than Δage in cores from the East Antarctic Plateau, and about one-third of Δage at other coastal sites that cover the last glacial period (Extended Data Fig. 1a)[23]. As the uncertainty in Δage is proportional to Δage itself, this leads to smaller dating uncertainties at WDC. Second, recent technological developments in coupling laser spectrometers with ice core continuous-flow analysis[34] have resulted in a WDC CH$_4$ record with the highest temporal resolution of all Antarctic ice cores so far. Extended Data Fig. 2c compares the WDC CH$_4$ record (grey, ~2-year sampling resolution) with the EDML

CH$_4$ record (orange, ~100-year sampling resolution). The estimated uncertainty in WDC Δage (ref. 23) is shown in Extended Data Fig. 1b–e.

Note that the transition between the layer-counted and CH$_4$-synchronized sections of the WD2014 chronology at 31.2 kyr does not influence our analysis of the bipolar seesaw phasing. In the stacking procedure we align each of the individual Antarctic events at the midpoint of the associated DO CH$_4$ transition. In doing so, we effectively synchronize all the WDC events to NGRIP $\delta^{18}O$, not just the events between 68 and 31.2 kyr. In performing the cross-correlation between NGRIP and WDC (Fig. 2a), we only analyse the CH$_4$-synchronized part of the record.

A dynamical version of the firn densification model[23,38] is used to calculate the Δage of the WD2014 chronology[23]. We furthermore calculated WDC Δage using an alternative firn densification model[23,39,40] (Extended Data Fig. 1b, blue curve). Using this alternative Δage scenario gives a NH lead time of 242 and 247 years for abrupt DO warming and DO cooling, respectively, in good agreement (within the 1σ uncertainty bound) with results presented in the text.

Identification of abrupt DO transitions. We use the midpoint of the abrupt transitions as time markers to align individual events. The method for finding the midpoint is identical to that used in ref. 23, where further details can be found. The method is shown in Extended Data Fig. 2a, b for DO 17 and 16 (we use a recently proposed DO nomenclature[41]). For each transition we define a pre-transition level and a post-transition level (horizontal orange markers), and use linear interpolation to find the depth at which the tracer of interest ($\delta^{18}O$ or CH$_4$) has completed 25%, 50% and 75% of the transition from the pre-transitional to post-transitional value. The 50% depth (red dots) is used to define the timing of the abrupt event; the 25% and 75% depths (blue dots) provide an uncertainty range on the midpoint[23].

It has been suggested that interhemispheric CH$_4$ synchronization is complicated by the fact that gas bubbles in ice cores have a gas age distribution, rather than a single age[42]. To investigate this effect at WDC, we construct a hypothetical gas age distribution (Extended Data Fig. 2d), using a (truncated) log-normal distribution as suggested elsewhere[42,43], in which we conservatively set $\sigma = 1$ and $\mu = \ln(50)$. The spectral width (or second moment) of this age distribution[44] is 24.3 years; for comparison, the gas-age spectral width for present-day WDC firn is only 4.8 years[45]. Applying this filter to an atmospheric ramp in CH$_4$ (Extended Data Fig. 2e) shifts the midpoint of the transition (as recorded in the ice core) forwards by 5 years. Because we have conservatively chosen the age distribution to be very wide, this 5-year shift should be considered an upper bound. The influence of the gas age distribution on the transition midpoint identification is very small in comparison with other sources of uncertainty, and is neglected in the remainder of the manuscript.

Stacking procedure. We define a time vector τ that runs from time $t = -1,200$ to $t = 1,200$ in 1-year increments. Next, for each of the DO/AIM events, we set the midpoint of the NGRIP $\delta^{18}O$ transition to $t = 0$, and use linear interpolation to sample the different records onto time vector τ. For each individual DO/AIM event this results in a record that spans from $-1,200$ to $+1,200$ years relative to the abrupt transition, at annual time steps; we shall refer to these as the contributory records. To align the WDC records we set the midpoint in the WDC CH$_4$ transition to $t = 56$ years, using the observation that the CH$_4$ transition slightly lags the Greenland $\delta^{18}O$ signal[23,24,46,47] (the uncertainty in this time lag is evaluated in the sensitivity study). With all the individual events synchronized and resampled to identical time spacing, we can average them to obtain the stacked record.

The spacing between DO events varies widely over time, ranging from several thousands of years to only hundreds of years. This means that in several cases the 2,400-year window we use contains more than a single event. We crop the contributory records whenever an adjacent event occurs within the sampling window (Extended Data Fig. 3). For widely spaced (that is, long-duration) events such as DO/AIM 12, no adjacent events fall within the time window, and no cropping is required (Extended Data Fig. 3a). For closely spaced events, such as DO/AIM 17.1, we need to crop neighbouring events (Extended Data Fig. 3b). The cropped parts of the contributory records are replaced with constant values that equal the boundary values (50-year averages) of the uncropped part of the record. The number of contributory records available as a function of time is shown in the bottom half of Extended Data Fig. 4. The shortest contributory record is the AIM 17.1 record, which is 700 years long.

Determining the breakpoint in the WDC $\delta^{18}O$ stack. We use an automated routine that is similar to the BREAKFIT algorithm[48], the main difference being the use of a second-order polynomial (rather than a linear) fit to the data, to account for the fact that the WDC $\delta^{18}O$ stack is curved rather than linear. The algorithm finds the breakpoint that minimizes the root mean square deviation (r.m.s.d.) between the stack (that is, the data) and the polynomial fit in the -600 to $+700$ years time interval (Extended Data Fig. 4); the interval was chosen

asymmetrically around zero to account for the fact that the breakpoint is always found at positive values. The routine operates as follows.

First, the user selects the range in which the algorithm is to look for the breakpoint, as well as a time step. For example, if the user selects 0–400 years at a 100-year time step, the algorithm will test the possibilities $t = 0, 100, 200, 300$ and 400 years. We shall refer to this one-dimensional array of values as the input vector.

Second, at each of the values in the input vector, the $\delta^{18}O$ stack is split into two pieces, and a second-order polynomial curve is fitted to data on each side separately. The two fitting curves are merged at their point of intersection. Note that the point of intersection can differ from the point where the data series were split in two.

Third, for each fit thus obtained, we calculate the r.m.s.d. to the stacked data over the -600 to $+700$ years time interval, and we select the best fit. The point of intersection of the two curves, rather than the point where the data series were split into two (that is, the values in the input vector), is considered the breakpoint in the $\delta^{18}O$ stack.

For most of the results in this work we let the algorithm search in the 0–400-year range, at a 2-year time step. The exception is the stacks of randomly selected DO events (Extended Data Fig. 6), for which we used a wider range (-100 to $+500$ years) to account for the larger spread in solutions. Note that the effective range of breakpoint detection is wider than the input range, because the point of intersection of the fitting curves is permitted to lie outside the range specified by the input vector. The fitting curves generated by the algorithm are plotted in dark orange lines in Fig. 2b, c and Extended Data Figs 4a, b, 6a–d and 7b, c; the breakpoint is indicated with an orange dot. The MATLAB code implementing the fitting procedure is provided in Supplementary Information.

The reason for using a quadratic rather than a linear fitting procedure is that Antarctic temperatures do not change in a linear fashion either before or after the abrupt transitions. As a test, we apply our fitting procedure to the WDC $\delta^{18}O$ stack (Fig. 2b), using either linear or quadratic fitting functions, in which we change the size of the data window from 300 years to 1,600 years. The window is applied symmetrically, meaning that equal numbers of years are used before and after the detected breakpoint (800 years on each side for a 1,600-year window). Data falling outside this window are ignored in the fitting procedure.

We find that the performance of the quadratic fitting procedure is superior in two important aspects (Extended Data Fig. 5). First, by using a quadratic fit, the breakpoint we detect is independent of the window size, which is not true for the linear fit. Second, the quadratic method provides a better approximation to the data, as expressed by the r.m.s.d. between the data and the fitting curve. At small window sizes of <700 years, both methods agree very well, and the fit to the data is comparable in terms of the r.m.s.d. This is due to the fact that at small window sizes the curvature of the time series is not very pronounced and is well approximated by a linear trend. We have also tested the BREAKFIT routine[48], which uses a linear fit on both sides of the breakpoint. We find a very good agreement between the breakpoints identified using the BREAKFIT routine and our linear fitting routine, with a mean difference of 0.2 years between them. Using a moving block bootstrap analysis[48], the BREAKFIT algorithm suggests a 2σ uncertainty of 50.8 years in breakpoint identification.

Monte Carlo sensitivity study. To assess how the different uncertainties in our method affect our conclusions, we perform a Monte Carlo sensitivity study. The largest source of uncertainty is the difference between gas age and ice age, or Δage (ref. 49). Δage determines the relative timing of the WDC CH_4 and $\delta^{18}O$ records, and thereby the phasing we infer for WDC AIM events relative to Greenland climate[2,3,50]. Increasing Δage shifts WDC ice ages towards older ages, thereby decreasing the NH lead time that we infer; vice versa, decreasing Δage will shift the ice ages younger, increasing the observed NH lead time. The Δage reconstruction we use here is obtained from a dynamical firn-densification model[40,47,51–54] constrained by $\delta^{15}N-N_2$ data, a proxy for past firn-column thickness[55]. The firn-densification model results are in good agreement[23] with the alternative Δdepth method developed recently[56]. For our current purpose we use an ensemble of 1,000 Δage histories (Extended Data Fig. 1b–e) generated by varying a suite of input parameters of the firn-densification model[23]. We furthermore use four different depth–age interpolation strategies[23,57], providing a total of $1,000 \times 4 = 4,000$ alternative WDC chronologies.

For each of these 4,000 potential WDC chronologies we further perform a Monte Carlo simulation (100 realizations) in which we perturb the alignment of the contributory records. The rationale behind the Monte Carlo study is that there is an uncertainty in the synchronization of the contributory records, and we want to assess how this uncertainty impacts our conclusions. In the Monte Carlo study, a random timing error is assigned to each of the contributory records, and thereby they are effectively shifted relative to one another in time. We distinguish between systematic and non-systematic errors, in which the former are correlated between

the different AIM events, and the latter are not. Sensitivity tests with the firn densification model show that Δage is a systematic error; if we have overestimated glacial Δage by, for example, 50 years, this will probably be true for all of the AIM events. We consider five independent sources of uncertainty in the synchronization procedure.

First, the midpoint detection in the abrupt NGRIP $\delta^{18}O$ transitions has an uncertainty; we shift the record by a value drawn from a Gaussian distribution of zero mean and a width that equals the midpoint uncertainty as defined above. This error is non-systematic.

Second, as in the first point above, but for the midpoint in the abrupt WDC CH_4 transition.

Third, although all four interpolation strategies synchronize the abrupt NH warming events between NGRIP and WDC, only two of them synchronize the NH cooling events[23]. For stacks of NH cooling we added an additional uncertainty drawn from a Gaussian distribution of zero mean and a width corresponding to the age difference between the cooling event as detected in NGRIP $\delta^{18}O$ and WDC CH_4. This error is non-systematic.

Fourth, a recent extensive analysis suggests a lag of CH_4 behind Greenland $\delta^{18}O$ of 56 ± 38 years (2σ)[24]. There are two types of uncertainty to consider: first, the CH_4 lag can differ between individual DO events (non-systematic error), and second, the central estimate of 56 years has an inherent uncertainty (systematic error). We here assume that both are equally large, and in the stacking procedure we assign both a 27-year systematic error and a 27-year non-systematic error to the CH_4 lag time. Both are drawn independently from a Gaussian distribution, and their combined uncertainty is therefore $(27^2 + 27^2)^{1/2} = 38$ years (that is, the 2σ uncertainty bound[24]).

Fifth, the breakpoint fitting procedure has an intrinsic uncertainty. We estimate this by using the moving-block bootstrap analysis of the BREAKFIT algorithm[48], which suggests a 1σ uncertainty of 25.4 years (Extended Data Fig. 5a). We therefore apply a 50.8-year (2σ) systematic uncertainty (systematic because it applies to the full $\delta^{18}O$ stack rather than to individual events).

Each contributory record is then individually shifted by an amount that corresponds to the sum of the randomly assigned non-systematic errors. After realigning the contributory records in this manner, the entire stack is shifted in time by an amount that corresponds to the sum of randomly assigned systematic errors. We detect the breakpoint by using the fitting algorithm. For each of the 4,000 chronologies we test 100 realizations of the Monte Carlo realignment, thus obtaining a total of 4×10^5 ($1,000 \times 4 \times 100$) estimates of the WDC $\delta^{18}O$ breakpoint. No statistical methods were used to predetermine the sample size used in the Monte Carlo analysis. The outcome of the sensitivity study is presented in the histograms of Fig. 2e, f and Extended Data Fig. 6e–h.

The non-systematic errors, because they are independent of each other, accumulate slowly when added in quadrature. The systematic errors (such as the Δage uncertainty or the breakpoint fitting procedure) therefore dominate the final uncertainty estimate. Running the Monte Carlo experiment with Δage uncertainties withheld gives a 2σ uncertainty bound on the interpolar phasing of 60 and 68 years for the DO onset and DO termination stacks, respectively; using only Δage uncertainties gives a 2σ uncertainty bound of 69 years. The Δage uncertainty is therefore as important as all other sources of uncertainty combined.

Code availability. The MATLAB code used to stack the records and perform the Monte Carlo analysis are provided in Supplementary Information.

Alternative stacks. To exclude the possibility that timing of the WDC $\delta^{18}O$ stack is dominated by a few prominent AIM events we perform the following tests.

First, we distinguish between major AIM events that are very prominent in the record, and minor events. The major AIM events are AIM 17, 14, 12, 8 and 4; all of these are preceded by NH Heinrich events. These large events have historically been called A-events, and are clearly visible even in low-resolution Antarctic cores (AIM 8, 12, 14 and 17 have been referred to in earlier work as A1, 2, 3 and 4, respectively)[2,3]. The minor events are the remaining ones (AIM 3, 5.1, 5.2, 6, 7, 9, 10, 11, 13, 15, 16 and 18). When we compare a stack consisting of only the major AIM events with a stack of only the minor AIM events, we find no statistically significant difference in their timing (Extended Data Fig. 6a, b). The NH lead time for the major and minor AIM stacks is 230 ± 93 and 211 ± 94 years, respectively (Extended Data Fig. 6e, f); the timing of the full stack (218 ± 92 years) is intermediate to these values, as expected (Fig. 2b).

In a second experiment we randomly selected eight DO/AIM events in the stacking procedure for each of the 4×10^5 realizations of the Monte Carlo experiment; one such realization is shown in Extended Data Fig. 6c, d for abrupt NH warming and cooling, respectively. The NH lead time we find in this experiment is 210 ± 113 years and 208 ± 108 years for abrupt NH warming and cooling, respectively (Extended Data Fig. 6g, h). These mean values differ by only a few years from those found for the full WDC $\delta^{18}O$ stack (Fig. 2). Note that the distribution width of NH lead times in this experiment is wider than the

distribution width found for the full $\delta^{18}O$ stacks. The reason for this may be threefold: first, a stack with fewer contributory records has a smaller signal-to-noise ratio, making the breakpoint detection less certain; second, if the NH lead time differs slightly between the AIM events, the average timing of a randomly selected subset may differ from the average timing of the full set; and third, in the Monte Carlo procedure the contribution of non-systematic uncertainties scales roughly in proportion to $\sqrt{N}$, again suggesting a narrower distribution for higher N (the number of events in the stack). The wider distribution found on decreasing the number of events in the stack therefore does not represent a more realistic uncertainty bound on the phasing of the bipolar seesaw.

Stacking the WDC Na record. Last, we investigate the phasing of the WDC sea-salt Na (ssNa) record relative to Greenland and Antarctic climate. The sea-ice surface is a major source of sea-salt aerosols to the atmosphere around Antarctica, and consequently ssNa has been interpreted as a proxy for the production and extent of regional sea ice[25,58,59]. Sea-salt Na is strongly anti-correlated with Antarctic climate on both orbital and millennial timescales[58,60].

Measured ssNa concentrations can be converted to a ssNa flux by multiplying with the accumulation rates at the site. On the East Antarctic plateau, the Na flux is expected to provide a better metric for atmospheric concentrations, because dry deposition dominates as a result of the very low accumulation rates[58]. Accumulation rates are much higher at WAIS Divide, however, and Greenland ice cores provide a better analogue than those on the East Antarctic plateau. It has been shown at Greenland Summit[61] that sea-salt concentrations in the ice reflect the atmospheric variations within 5%, whereas the sea-salt flux underestimates the atmospheric variations by about 50%. Thus, the observed variations in WDC ssNa concentrations should to first order reflect variations in the atmospheric loading.

We investigate the relative timing of the ssNa record with the same tools as those that we applied to the WDC $\delta^{18}O$ record. An evaluation of the lagged correlation between NGRIP $\delta^{18}O$ and the time derivative of the various WDC records is shown in Extended Data Fig. 7a. For ssNa we obtain a centennial-scale NH lead time, comparable to that observed for WDC $\delta^{18}O$. Stacking of the WDC impurity records yields the same picture (Extended Data Fig. 7b, c); the breakpoint in the ssNa stacks coincides with the breakpoint in the WDC $\delta^{18}O$ stack, and lags the abrupt NH transition by roughly two centuries. Note that because $\delta^{18}O$ and ssNa are both measured in the ice phase, the uncertainty that we estimate for the timing of the $\delta^{18}O$ breakpoint (2σ of about 90 years) also holds for the impurity records.

The synchronicity of WDC ssNa and $\delta^{18}O$ variations suggests that Antarctic climate and sea-ice extent are closely linked on centennial, or perhaps even sub-centennial timescales. This may reflect a common forcing and/or a feedback between Southern Ocean surface temperatures and sea-ice extent. The sea-ice extent may furthermore be important in forcing Antarctic temperatures (and $\delta^{18}O$)[6,62], particularly at the WAIS Divide site, where the marine influence is stronger than on the East Antarctic plateau[17].

31. Steig, E. J. et al. Recent climate and ice-sheet changes in West Antarctica compared with the past 2,000 years. Nature Geosci. **6**, 372–375 (2013).
32. Sowers, T. et al. An interlaboratory comparison of techniques for extracting and analyzing trapped gases in ice cores. J. Geophys. Res. **102**, 26527–26538 (1997).
33. Mitchell, L. E., Brook, E. J., Sowers, T., McConnell, J. R. & Taylor, K. Multidecadal variability of atmospheric methane, 1000–1800 C.E. J. Geophys. Res. **116**, G02007 (2011).
34. Stowasser, C. et al. Continuous measurements of methane mixing ratios from ice cores. Atmos. Meas. Tech. **5**, 999–1013 (2012).
35. Rhodes, R. H. et al. Continuous methane measurements from a late Holocene Greenland ice core: atmospheric and in-situ signals. Earth Planet. Sci. Lett. **368**, 9–19 (2013).
36. Dlugokencky, E. et al. Conversion of NOAA atmospheric dry air CH_4 mole fractions to a gravimetrically prepared standard scale. J. Geophys. Res. **110**, D18306 (2005).
37. Sigl, M. et al. A new bipolar ice core record of volcanism from WAIS Divide and NEEM and implications for climate forcing of the last 2000 years. J. Geophys. Res. Atmos. **118**, 1151–1169 (2013).
38. Herron, M. M. & Langway, C. C. Firn densification: an empirical model. J. Glaciol. **25**, 373–385 (1980).
39. Arnaud, L., Barnola, J. M. & Duval, P. in Physics of Ice Core Records (ed. Hondoh, T.) 285–305 (Hokkaido Univ. Press, 2000).
40. Goujon, C., Barnola, J. M. & Ritz, C. Modeling the densification of polar firn including heat diffusion: application to close-off characteristics and gas isotopic fractionation for Antarctica and Greenland sites. J. Geophys. Res. **108** (D24), 4792 (2003).
41. Rasmussen, S. O. et al. A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. Quat. Sci. Rev. **106**, 14–28 (2014).
42. Köhler, P. Rapid changes in ice core gas records. Part 1. On the accuracy of methane synchronisation of ice cores. Clim. Past Discuss. **6**, 1453–1471 (2010).
43. Köhler, P., Fischer, H. & Schmitt, J. Atmospheric $\delta^{13}CO_2$ and its relation to pCO_2 and deep ocean $\delta^{13}C$ during the late Pleistocene. Paleoceanography **25**, PA1213 (2010).
44. Trudinger, C. M. et al. Reconstructing atmospheric histories from measurements of air composition in firn. J. Geophys. Res. **107** (D24) 4780 (2002).
45. Battle, M. O. et al. Controls on the movement and composition of firn air at the West Antarctic Ice Sheet Divide. Atmos. Chem. Phys. **11**, 11007–11021 (2011).
46. Huber, C. et al. Isotope calibrated Greenland temperature record over Marine Isotope Stage 3 and its relation to CH_4. Earth Planet. Sci. Lett. **243**, 504–519 (2006).
47. Rasmussen, S. O. et al. A first chronology for the North Greenland Eemian Ice Drilling (NEEM) ice core. Clim. Past **9**, 2713–2730 (2013).
48. Mudelsee, M. Break function regression. Eur. Phys. J. Spec. Top. **174**, 49–63 (2009).
49. Schwander, J. & Stauffer, B. Age difference between polar ice and the air trapped in its bubbles. Nature **311**, 45–47 (1984).
50. Capron, E. et al. Synchronising EDML and NorthGRIP ice cores using $\delta^{18}O$ of atmospheric oxygen ($\delta^{18}O_{atm}$) and CH_4 measurements over MIS5 (80–123 kyr). Quat. Sci. Rev. **29**, 222–234 (2010).
51. Barnola, J. M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. CO_2–climate relationship as deduced from the Vostok ice core: a re-examination based on new measurements and on a re-evaluation of the air dating. Tellus **43**, 83–90 (1991).
52. Schwander, J. et al. Age scale of the air in the summit ice: implication for glacial–interglacial temperature change. J. Geophys. Res. **102**, 19483–19493 (1997).
53. Seierstad, I. et al. Consistently dated records from the Greenland GRIP, GISP2 and NGRIP ice cores for the past 104 ka reveal regional millennial-scale isotope gradients with possible Heinrich Event imprint. Quat. Sci. Rev. **106**, 29–46 (2014).
54. Buizert, C. et al. Greenland temperature response to climate forcing during the last deglaciation. Science **345**, 1177–1180 (2014).
55. Sowers, T., Bender, M., Raynaud, D. & Korotkevich, Y. S. $\delta^{15}N$ of N_2 in air trapped in polar ice: a tracer of gas transport in the firn and a possible constraint on ice age–gas age differences. J. Geophys. Res. **97**, 15683–15697 (1992).
56. Parrenin, F. et al. On the gas–ice depth difference (Δdepth) along the EPICA Dome C ice core. Clim. Past **8**, 1239–1255 (2012).
57. Fudge, T. J., Waddington, E. D., Conway, H., Lundin, J. M. D. & Taylor, K. Interpolation methods for Antarctic ice-core timescales: application to Byrd, Siple Dome and Law Dome ice cores. Clim. Past Discuss. **10**, 65–104 (2014).
58. Fischer, H., Siggaard-Andersen, M.-L., Ruth, U., Röthlisberger, R. & Wolff, E. Glacial/interglacial changes in mineral dust and sea-salt records in polar ice cores: sources, transport, and deposition. Rev. Geophys. **45**, RG1002 (2007).
59. de Vernal, A., Gersonde, R., Goosse, H., Seidenkrantz, M.-S. & Wolff, E. W. Sea ice in the paleoclimate system: the challenge of reconstructing sea ice from proxies—an introduction. Quat. Sci. Rev. **79**, 1–8 (2013).
60. Wolff, E. W. et al. Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles. Nature **440**, 491–496 (2006).
61. Alley, R. et al. Changes in continental and sea-salt atmospheric loadings in central Greenland during the most recent deglaciation: model-based estimates. J. Glaciol. **41**, 503–514 (1995).
62. Noone, D. & Simmonds, I. Sea ice control of water isotope transport to Antarctica and implications for ice core interpretation. J. Geophys. Res. **109**, D07105 (2004).
63. Veres, D. et al. The Antarctic ice core chronology (AICC2012): an optimized multi-parameter and multi-site dating approach for the last 120 thousand years. Clim. Past **9**, 1733–1748 (2013).
64. Bazin, L. et al. An optimized multi-proxy, multi-site Antarctic ice and gas orbital chronology (AICC2012): 120–800 ka. Clim. Past **9**, 1715–1731 (2013).
65. Kawamura, K. et al. Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. Nature **448**, 912–916 (2007).
66. Schilt, A. et al. Atmospheric nitrous oxide during the last 140,000 years. Earth Planet. Sci. Lett. **300**, 33–43 (2010).
67. Köhler, P., Knorr, G., Buiron, D., Lourantou, A. & Chappellaz, J. Abrupt rise in atmospheric CO_2 at the onset of the Bølling/Allerød: in-situ ice core data versus true atmospheric signals. Clim. Past **7**, 473–486 (2011).
68. Rosen, J. L. et al. An ice core record of near-synchronous global climate changes at the Bølling transition. Nature Geosci. **7**, 459–463 (2014).
69. Köhler, P., Knorr, G. & Bard, E. Permafrost thawing as a possible source of abrupt carbon release at the onset of the Bølling/Allerød. Nature Comm **5**, 5520 (2014).
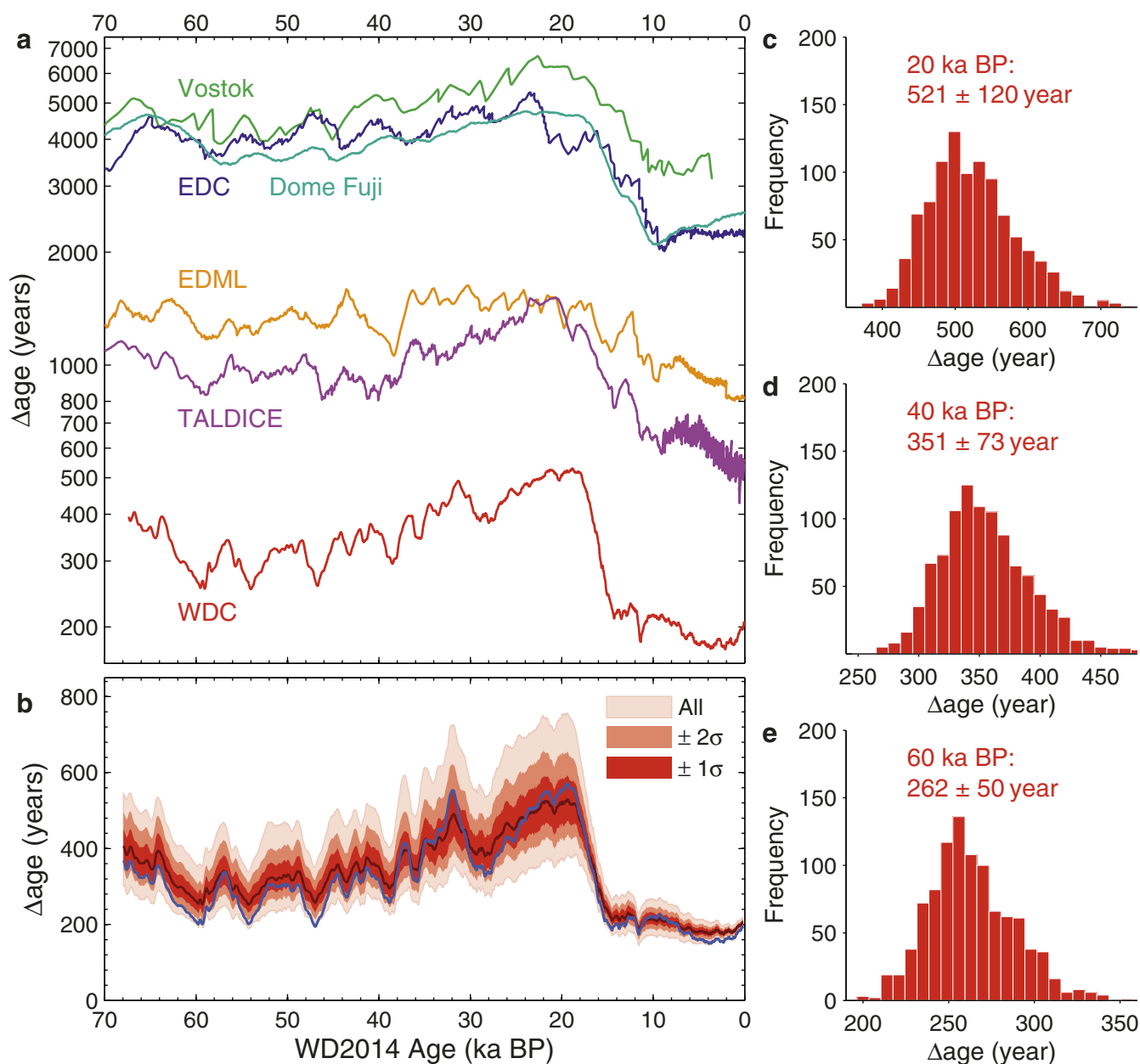
Extended Data Figure 1 | Difference between gas age and ice age (Δage) at WAIS Divide. a, Comparison of WDC Δage with other Antarctic cores. Ice core abbreviations: EDC, EPICA Dome Concordia; EDML, EPICA Dronning Maud Land; TALDICE, Talos Dome; WDC, WAIS Divide. Δage values are taken from refs 23, 63–65. The vertical axis is on a logarithmic scale. **b,** Δage uncertainty bounds obtained from an ensemble of 1,000 alternative Δage scenarios; details are given elsewhere[23]. A Δage scenario obtained with an alternative densification model (ref. 39 instead of ref. 38) is shown in blue. **c–e,** Histograms of the 1,000 Δage scenarios at 20 kyr BP (**c**), 40 kyr BP (**d**) and 60 kyr BP (**e**); stated values give the distribution mean ± the 2σ standard deviation.
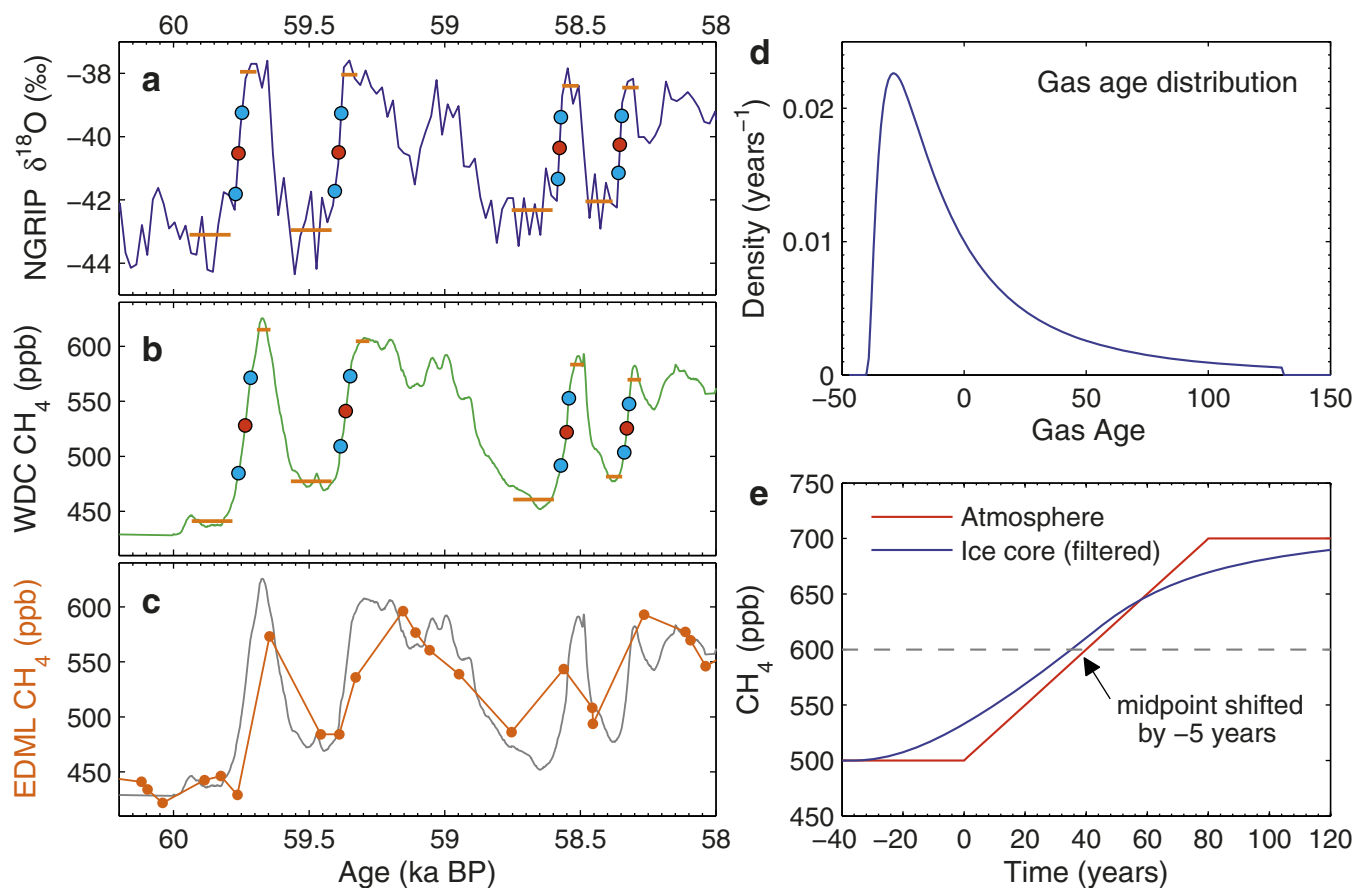
Extended Data Figure 2 | Determining the timing of the abrupt DO transitions. **a**, **b**, DO 17.2, 17.1, 16.2 and 16.1 (from oldest to youngest[41]) as recorded in NGRIP $\delta^{18}O$ (**a**) and WDC CH_4 (**b**). Horizontal orange bars denote pre-transition and post-transition levels; the transition midpoint (50% of signal amplitude) is indicated by a red dot; the 25% and 75% signal amplitude markers are indicated with blue dots. **c**, Comparison of WDC CH_4 (grey) with EDML CH_4 (orange)[3,50,66]. **d**, Hypothetical gas-age distribution for WDC due to firn densification and gradual bubble closure, using a truncated log-normal distribution[67]. **e**, Shift in transition midpoint induced by filtering of the atmospheric record in the firn column.
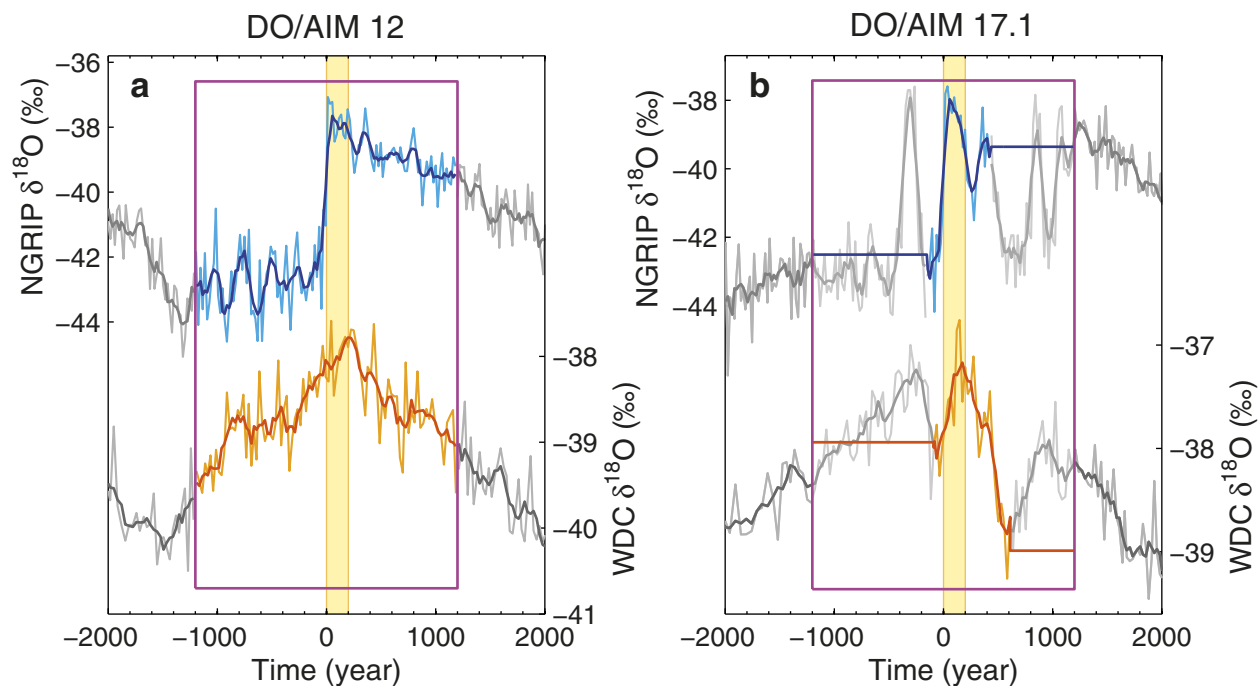
Extended Data Figure 3 | Cropping of individual records in the stack to prevent overlap of events. a, DO/AIM 12, where no cropping is needed. **b**, DO/AIM 17.1, where the most cropping is needed. Full time series with five-point running average are plotted in grey, and the contributory records are plotted in blue and orange for NGRIP and WDC, respectively. The yellow vertical shading bar in background shows the NH lead time (200 years); the purple rectangle gives the −1,200 to +1,200 time window.
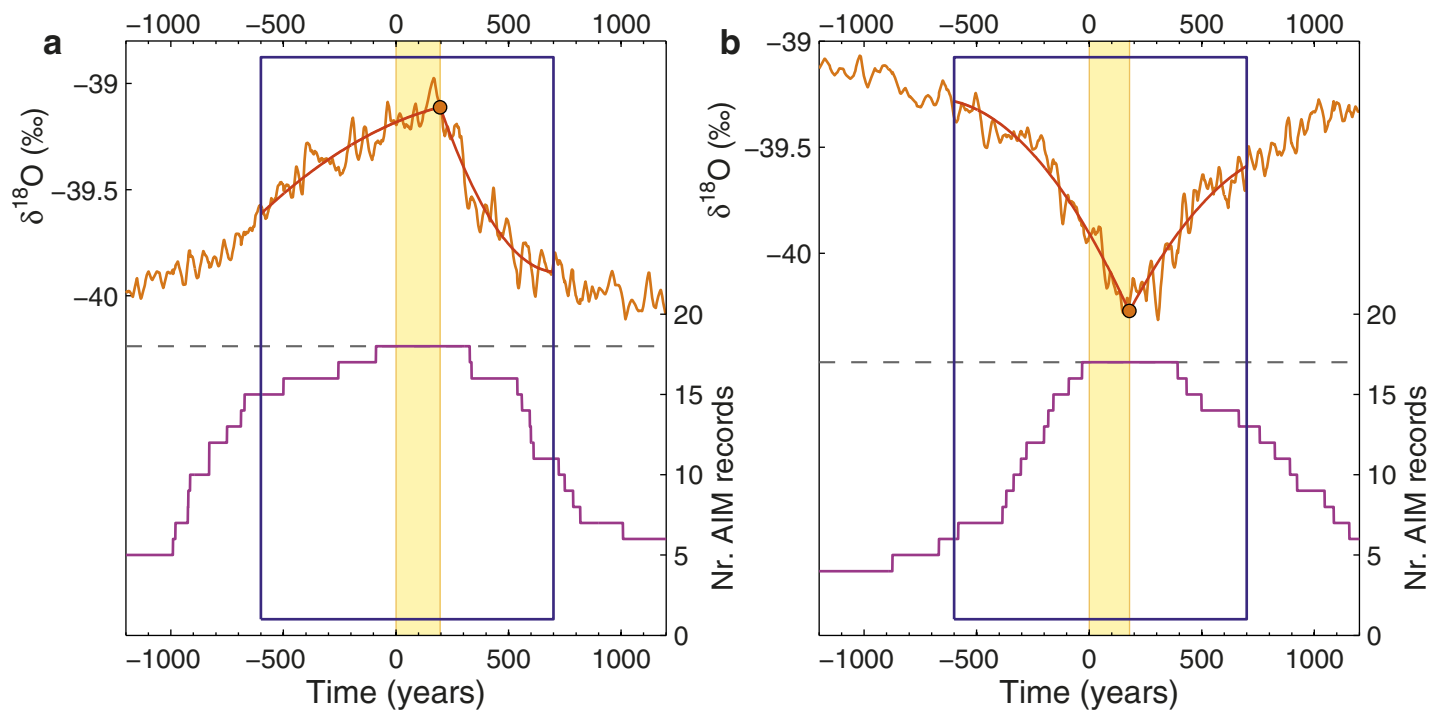
Extended Data Figure 4 | Number of records and fitting procedure. Number of contributory records to the WDC $\delta^{18}O$ stacks for abrupt NH warming (interstadial onset) (**a**) and for abrupt NH cooling (interstadial termination) (**b**). Blue rectangles indicate the time window over which the fitting procedure evaluates the fit to the data (-600 to $+700$ years); shaded vertical yellow bars show NH lead time.
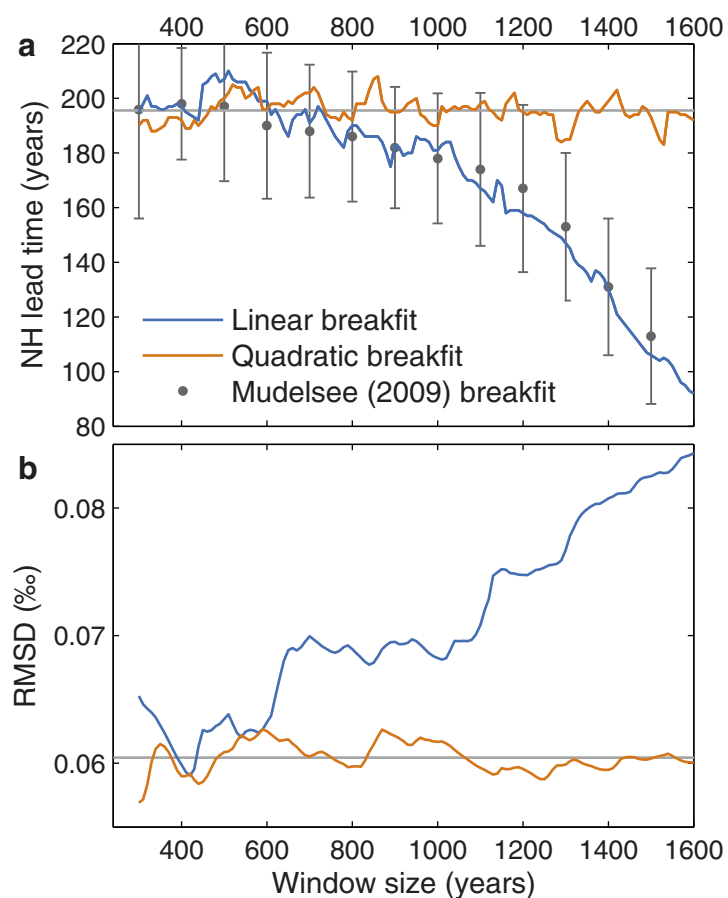
Extended Data Figure 5 | Evaluating the performance of the breakpoint detection algorithm. a, Breakpoint detection as a function of data window size using both linear (blue) and quadratic (orange) functions, compared with the BREAKFIT algorithm[48] (grey dots with 1σ error bars). The data window is applied symmetrically, meaning that equal numbers of years (half the window size) are used before and after the detected breakpoint. Data falling outside this window are ignored in the fitting procedure. **b**, Root mean square deviation between the WDC $\delta^{18}O$ stack and the fitting curve.
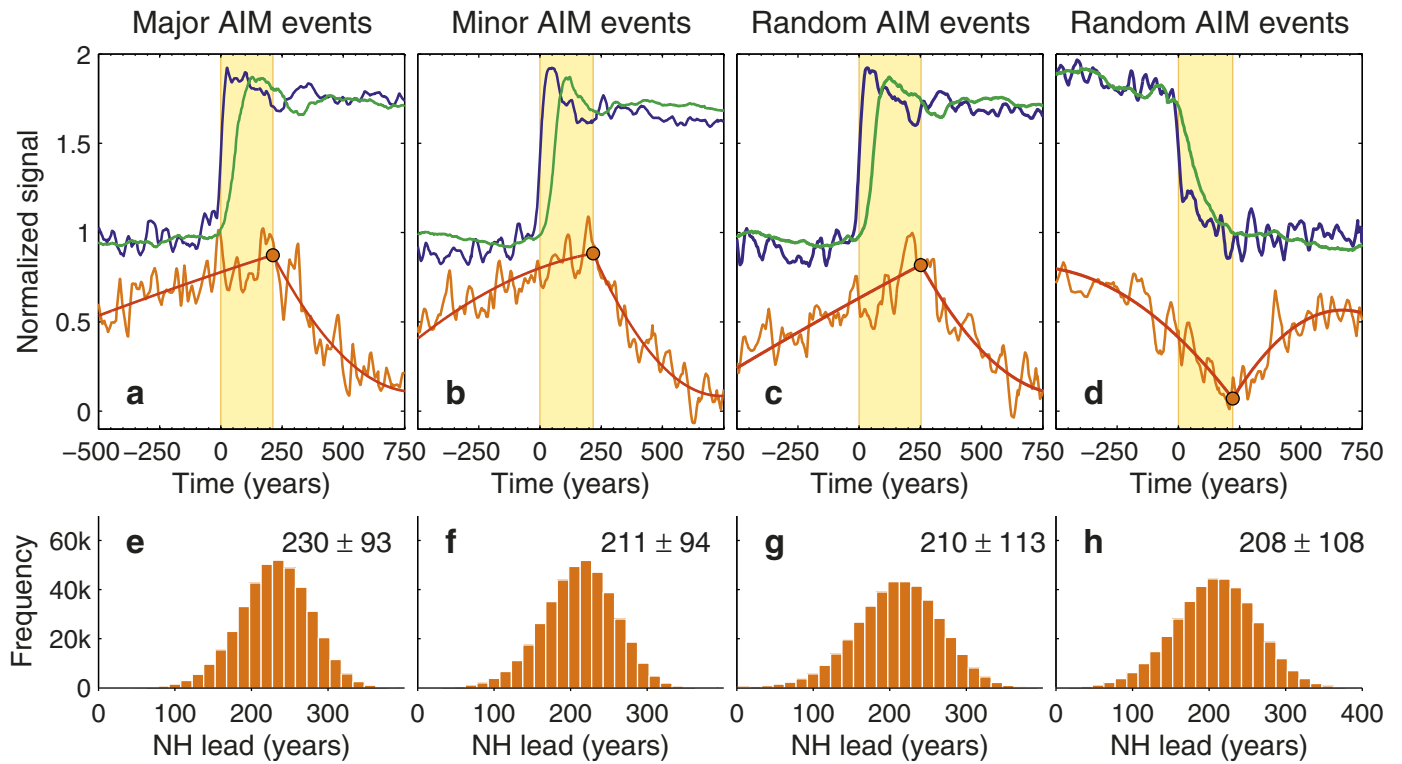
Extended Data Figure 6 | Alternative stacking of AIM events. a, Stack of NGRIP δ^{18}O (blue), WDC CH$_4$ (green) and WDC δ^{18}O (orange with fit) for just the major AIM events (4, 8, 12, 14 and 17), aligned at the abrupt NH warming. **b**, As in **a**, but for only the minor AIM events (3, 5.1, 5.2, 6, 7, 9, 10, 11, 13, 15, 16 and 18). **c**, As in **a**, but for eight randomly selected DO/AIM events. **d**, As in **c**, but aligned at the abrupt NH cooling. Events are averaged with their original amplitudes and normalized after stacking for convenience of visualization. **e–h**, Histograms of NH lead time associated with **a–d**, respectively, generated by binning the 4×10^5 solutions from the sensitivity study. The distribution mean and 2σ uncertainty bounds are listed in the panels. Shaded vertical yellow bars (upper panels) show NH lead time.
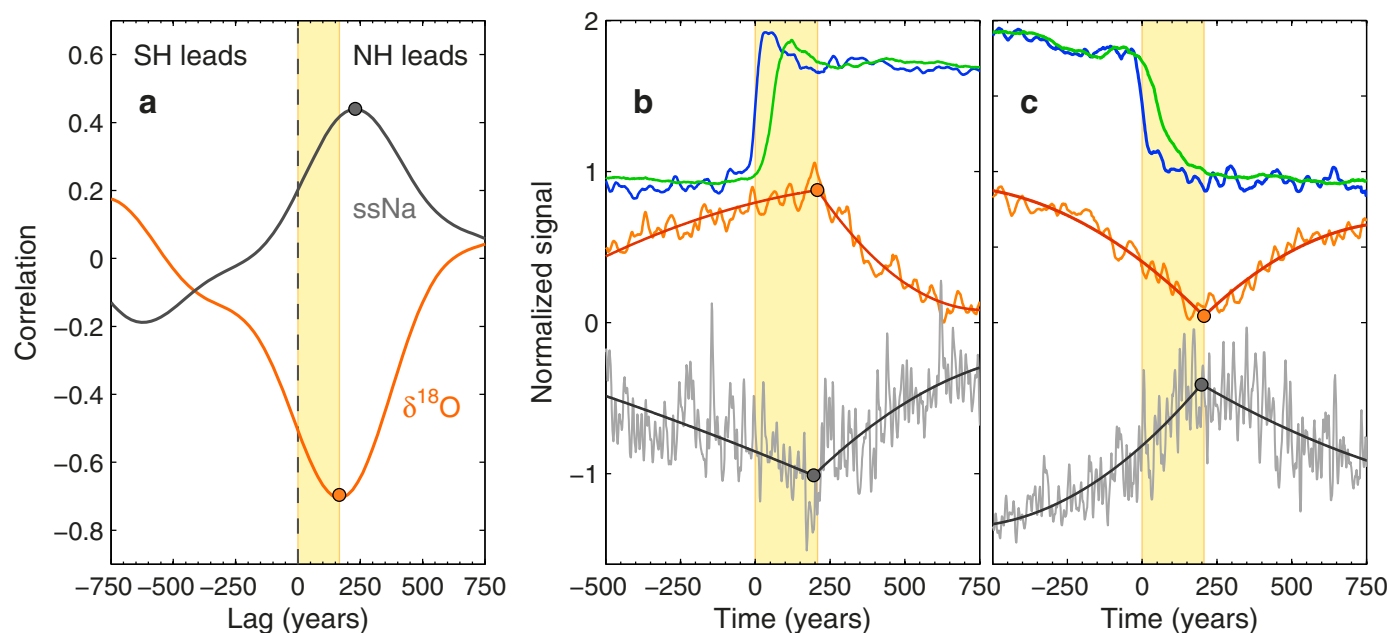
Extended Data Figure 7 | Timing of sea-salt sodium. **a**, Lagged correlation between NGRIP δ^{18}O and the time derivative of WDC δ^{18}O (orange), and between NGRIP δ^{18}O and the time derivative of WDC ssNa (grey). The dots indicate the maximum (anti-)correlation at 167-year and 229-year NH lead for WDC δ^{18}O and ssNa, respectively. A fourth-order Butterworth bandpass filter with a 500–10,000-year window is applied to the time series to isolate the millennial-scale variability. **b**, DO3–18 stack of NGRIP δ^{18}O (blue), WDC CH$_4$ (green), WDC δ^{18}O (orange) and WDC ssNa (grey), aligned at the midpoint of the DO warming signal. The estimated breakpoint in the stacks (dots) occurs at $t = 218$ and 195 years for WDC δ^{18}O and ssNa, respectively. **c**, As in **b**, but for the abrupt NH cooling events, with the estimated breakpoint at $t = 208$ and 199 years for WDC δ^{18}O and ssNa, respectively. Shaded vertical yellow bars show NH lead times.

Extended Data Table 1 | Phasing of the bipolar seesaw during the last deglaciation

Transition	WDC δ^{18}O breakpoint	WDC CH$_4$ Transition	Δage	Greenland – CH$_4$ phasing	Lead of NH climate
OD - B/A	14,350 ± 110	14,576 ± 29	228 ± 62	30 ± 32	256 ± 133
B/A – YD	12,778 ± 80	12,764 ± 53	218 ± 68	30 ± 32	16 ± 122
		12,881 ± 30†	216 ± 67	30 ± 32	133 ± 113†
YD – Holoc.	11,842 ± 90‡	11,542 ± 33	195 ± 52	30 ± 32	−270 ± 114‡

For the deglaciation we use a Greenland-CH$_4$ phasing that is the average of the 56 ± 38-year CH$_4$ lag observed for the glacial period[24,46], and the near-synchronous phasing (4.5 ± 24-year CH$_4$ lag) observed for the Bølling transition[68]. All stated errors represent 2σ uncertainty bounds; uncertainty in the WDC δ^{18}O breakpoint is determined using the BREAKFIT algorithm[48]; uncertainty in the CH$_4$ transition is defined as the 25–75% range of the CH$_4$ transition (Extended Data Fig. 2); Δage uncertainty is found by using a firn densification model sensitivity study[23] (Extended Data Fig. 1); uncertainty in the Greenland-CH$_4$ phasing is taken to be the root mean square of the uncertainties in the cited studies[24,68]; the uncertainty in the NH climate is the root sum square of the uncertainties in the preceding columns. The potential role of CO_2 in delaying Antarctic cooling at the Bølling onset is discussed in ref. 69.
† This estimate uses the onset, rather than the midpoint, of the Bølling–Allerød to Younger Dryas transition in WDC CH$_4$ in determining the phasing.
‡ This estimate may be unreliable because of a double peak in WDC δ^{18}O at about this time that may reflect local climate, as well as a unique and pronounced accumulation anomaly starting at about 11.9 kyr BP.

LETTER

Isotopic evidence for biological nitrogen fixation by molybdenum–nitrogenase from 3.2 Gyr

Eva E. Stüeken[1], Roger Buick[1], Bradley M. Guy[2]† & Matthew C. Koehler[1]

Nitrogen is an essential nutrient for all organisms that must have been available since the origin of life. Abiotic processes including hydrothermal reduction[1], photochemical reactions[2], or lightning discharge[3] could have converted atmospheric N_2 into assimilable NH_4^+, HCN, or NO_x species, collectively termed fixed nitrogen. But these sources may have been small on the early Earth, severely limiting the size of the primordial biosphere[4]. The evolution of the nitrogen-fixing enzyme nitrogenase, which reduces atmospheric N_2 to organic NH_4^+, thus represented a major breakthrough in the radiation of life, but its timing is uncertain[5,6]. Here we present nitrogen isotope ratios with a mean of 0.0 ± 1.2‰ from marine and fluvial sedimentary rocks of prehnite–pumpellyite to greenschist metamorphic grade between 3.2 and 2.75 billion years ago. These data cannot readily be explained by abiotic processes and therefore suggest biological nitrogen fixation, most probably using molybdenum-based nitrogenase as opposed to other variants that impart significant negative fractionations[7]. Our data place a minimum age constraint of 3.2 billion years on the origin of biological nitrogen fixation and suggest that molybdenum was bioavailable in the mid-Archaean ocean long before the Great Oxidation Event.

Nitrogenase is present in approximately 15% of all sequenced prokaryotic genomes, including several phyla of Bacteria and one phylum of Archaea, but it is absent in Eukaryota[8]. The most common form (Nif) contains molybdenum in the catalytically active site, but a subset of diazotrophic (N_2-fixing) organisms produce 'alternative nitrogenases' containing V (Vnf) or Fe (Anf)[9] instead. It was once thought that Vnf and Anf evolved before Nif, possibly in the Archaean era when seawater was enriched in Fe and depleted in Mo[6,10], but newer phylogenetic evidence indicates that Nif is ancestral to the alternative isozymes[5,9] and may itself have been preceded by a 'proto-nitrogenase' enzyme with low efficiency and substrate specificity[5,9]. Protein substitution rates[5], combined with the view that marine Mo was scarce before the Great Oxidation Event (GOE)[11], have been interpreted as indicating a relatively late origin of Mo-nitrogenase around 1.5–2.2 billion years (Gyr) ago[5]. However, other Mo-based enzymes, in particular the molybdopterin cofactor, have been traced back phylogenetically to the last universal common ancestor[12], implying that some Mo was bioavailable in the early Archaean. Furthermore, abiotic nitrogen sources may have become limiting long before the GOE[4], making it conceivable that Mo-based nitrogenase is more ancient than models suggest.

Nitrogen isotopes of sedimentary rocks can clarify this issue because metabolic nitrogen pathways exert characteristic fractionations that can be preserved in kerogen over geological time. Moderately positive nitrogen isotope ratios ($\delta^{15}N = [(^{15}N/^{14}N)_{sample}/(^{15}N/^{14}N)_{standard}] - 1$) from the late Archaean (2.7–2.5 Gyr) indicate an aerobic nitrogen cycle involving nitrification and denitrification[13,14]. It is widely assumed that before about 2.7 Gyr, the nitrogen cycle was anaerobic[15], consisting only of nitrogen fixation and ammonification. As the latter has negligible isotopic
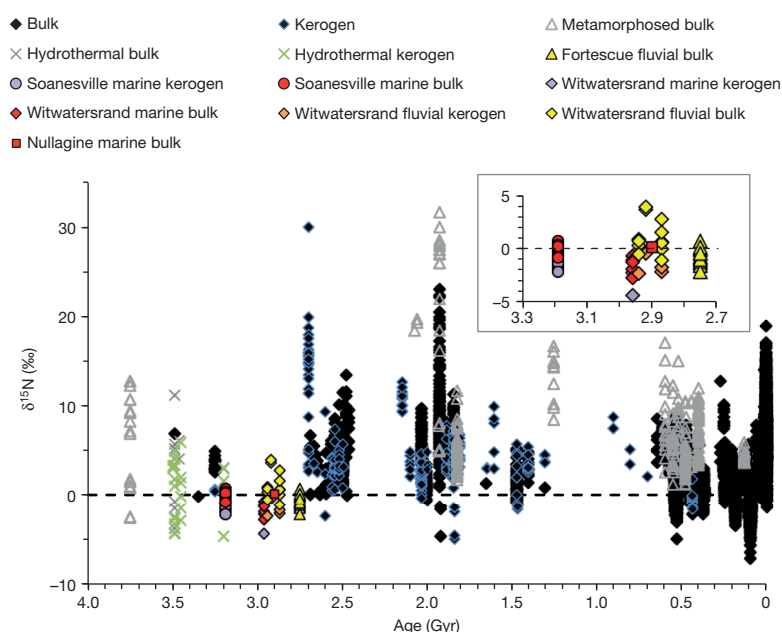
Figure 1 | Sedimentary $\delta^{15}N$ through time. All points labelled 'Soanesville', 'Witwatersrand', 'Nullagine', and 'Fortescue' are from this study, all others are compiled from the literature. 'Hydrothermal' refers to samples from hydrothermal veins; 'Metamorphosed', metamorphosed above greenschist facies; 'Witwatersrand', Witwatersrand Supergroup. Insert shows data from this study at higher resolution; note that Soanesville Group data at 3.2 Gyr represent 20 superimposed points. See Supplementary Information for references.

[1]Department of Earth & Space Sciences and Astrobiology Program, University of Washington, Seattle, Washington 98195-1310, USA. [2]Department of Geology, University of Johannesburg, Auckland Park 2006, South Africa. †Present address: Mineral Services, SGS South Africa, 58 Melville Street, Booysens, Johannesburg 2091, South Africa.

Table 1 | Sample localities and depositional environments

Unit	Age (Gyr)	Depositional setting	MG	n	$\delta^{15}N_{bulk}$ (‰)	TN (p.p.m.)	$\delta^{13}C_{org}$ (‰)	TOC (%)
Soanesville Group	3.19	Upper continental slope	PP	20	+0.1 ± 0.5	247 ± 146	−31.0 ± 0.3	1.3 ± 0.5
Nullagine Group	2.98	Outer shelf	GR	1*	+0.1 ± 0.4	42 ± 3	−40.7 ± 0.0	0.2 ± 0.0
Witwatersrand	2.92	Fluvial	GR	8	+0.9 ± 1.9	65 ± 63	−41.8 ± 7.1	0.2 ± 0.2
Supergroup	2.87	Shallow marine, near rivers	GR	5	+1.2 ± 1.0	47 ± 18	−37.8 ± 1.2	0.5 ± 0.2
	2.96	Marine shelf	GR	5	−1.6 ± 0.8	105 ± 48	−28.2 ± 0.9	0.5 ± 0.5
Fortescue Group	2.76	Fluvio-lacustrine	PP	13	−0.7 ± 0.8	22 ± 7	−36.0 ± 5.3	0.1 ± 0.1

MG, metamorphic grade: PP, prehnite–pumpellyite; GR, greenschist. n, Number of samples. Isotopic uncertainties are reported as 1σ.
* For the Nullagine sample, the reported standard deviation is equal to the analytical uncertainty of triplicate analyses.

effects ($\epsilon = 1$‰)[16], the isotopic signature of biological nitrogen fixation should be discernible in the earlier Archaean rock record. Nif can impart fractionations down to −4‰ in fixed nitrogen relative to atmospheric N_2, as seen in some thermophilic organisms, but most marine observations are in the range −2‰ to +1‰ (average −1‰)[7,17,18]. The less efficient alternative nitrogenases Vnf and Anf fractionate nitrogen isotopes by −6‰ to −8‰ (ref. 7). The fractionation imparted by potential proto-nitrogenases is unknown, but given their presumed low efficiency and serendipitous filling of their catalytic sites by various metals, including Fe or V (ref. 9), fractionations may have been relatively large, close to values observed for Vnf and Anf, and probably quite variable. The composition of atmospheric N_2 has evidently been between −2‰ and +2‰ since the early Archaean[19]. Hence, if $\delta^{15}N$ values of Archaean sediments fall consistently within a narrow range, then they may reveal if nitrogenase had evolved and which isozyme was dominant. However, published $\delta^{15}N$ data from early and mid-Archaean rocks are widely scattered (−3.8‰ to +12.7‰; Fig. 1), partly because of metamorphic alteration to above greenschist facies, metasomatism by ammoniated brines, hydrothermal input during deposition, or high analytical uncertainty (ref. 15 and Supplementary Information 1). Existing data are thus inconclusive as to the presence of microbial nitrogen fixation.

To place a better constraint on the origin of nitrogenase, we analysed nitrogen and carbon isotope ratios in mudrocks of prehnite–pumpellyite to greenschist metamorphic grade from marine and non-marine environments between 3.2 and 2.75 Gyr (Table 1 and Supplementary Information 2). The samples showed no evidence of post-depositional fluid alteration. Contributions of detrital nitrogen were minor and limited to fluvial samples (Supplementary Information 3.1 and Extended Data Fig. 1). We analysed decarbonated whole-rock samples and a few kerogen isolates by elemental analyser–isotope ratio mass spectrometry (ref. 20 and Methods). Unlike cherts, which were commonly used in previous studies, mudrocks are relatively kerogen-rich and widespread, making them ideal for obtaining accurate data over long timescales. Phyllosilicate minerals can accumulate NH_4^+ released from kerogen during diagenesis. Therefore, whole-rock samples containing both organic nitrogen and mineral-bound ammonium should come closest to preserving the original isotopic composition of biomass (Supplementary Information 3.1), if infiltration of fluids from other sources can be ruled out. To detect possible alteration, as well as to quantify the availability of Mo, V, and Fe in the depositional environments, we determined major and minor element abundances by inductively coupled plasma–atomic emission spectrometry and inductively coupled plasma–mass spectrometry (Supplementary Tables A3 and A4).

Our results for marine samples ($n = 26$) showed a narrow range (−2.8‰ to +0.7‰; Fig. 2) for mid-Archaean $\delta^{15}N_{bulk}$ with an average of −0.2 ± 0.9‰. Most notably, 20 samples from the 3.2 Gyr Soanesville Group, collected from separate drill-cores spanning several hundred metres of stratigraphy, almost overlapped, ranging from −0.9‰ to +0.7‰ (mean = +0.1 ± 0.5‰). Fluvial samples from the Fortescue Group and the Witwatersrand Supergroup ($n = 26$) had a similar mean value (+0.2 ± 1.5‰); only four samples were heavier than +1‰. We included marine Witwatersrand sediments deposited proximal to river inflows in the fluvial category, because they were probably influenced by freshwater runoff. Their organic carbon isotopes (Table 1) were similar to the fluvial facies, supporting this reclassification. The four heavier $\delta^{15}N_{bulk}$ values in fluvial samples from the Witwatersrand basin may reflect local

redox processes, consistent with sulphur isotope data indicating mild oxidative weathering on land at this time[21]. Apart from these four points, all our data fell within the range of Mo-based nitrogen fixation (Fig. 2). This is also true if $\delta^{15}N$ values from the Witwatersrand Supergroup and Nullagine Group increased by 1–2‰ during greenschist metamorphism[15]. The 17 kerogen isolates were lighter than whole rocks (by −1.4‰ to −4.1‰; Extended Data Table 1), consistent with diagenetic alteration under anoxic conditions (Supplementary Information 3.1). Nitrogen isotopic composition was independent of total organic carbon content (TOC; Fig. 2) and hence was probably not influenced by productivity, because sedimentation rates, the other control on TOC content, are unlikely to be highly variable in similar environments such as within the Soanesville Group. Elemental abundances in the Soanesville Group and the Witwatersrand Supergroup, including concentrations of Mo and V, were generally close to crustal averages (Fig. 3 and Extended Data Fig. 2). This pattern argues against significant metasomatic overprinting.

The narrow range in $\delta^{15}N_{bulk}$ close to the presumed composition of the atmosphere is inconsistent with an excess supply of fixed nitrogen from abiotic sources, because partial assimilation of either NH_4^+ or NO_x species would have resulted in substantial fractionations of greater than 10‰ (ref. 16). External sources of fixed nitrogen can therefore only be reconciled with our data if assimilation was quantitative, and if those fixed nitrogen species had an initial $\delta^{15}N$ composition near 0‰. It is unlikely that abiotically fixed nitrogen from the atmosphere would fulfil this condition. Organic nitrogen compounds formed during electric discharge experiments are fractionated relative to N_2 by −15‰ to −25‰, and photodissociation of N_2 by ultraviolet light leads to even

Legend:
- Soanesville marine bulk
- Witwatersrand marine bulk
- Witwatersrand fluvial bulk
- Fortescue fluvial bulk
- Soanvesville marine kerogen
- Witwatersrand marine kerogen
- Witwatersrand fluvial kerogen
- Nullagine marine bulk
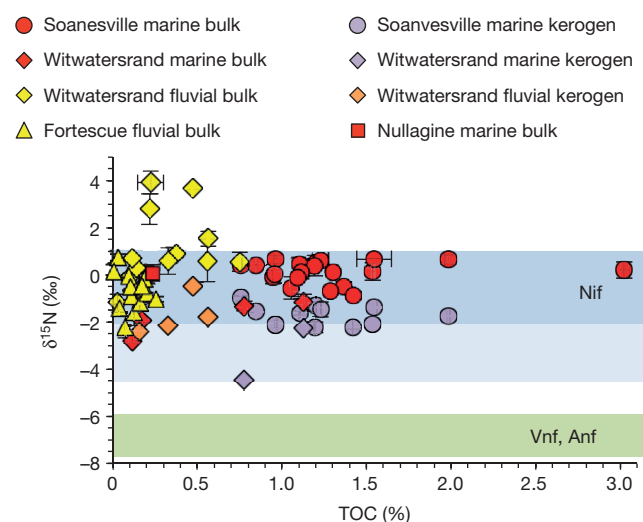
Figure 2 | Whole-rock $\delta^{15}N$ versus TOC. Symbols are as in Fig. 1. Error bars are 1σ of two or three replicate analyses of the same decarbonated sample aliquot. Dark blue shading, range of $\delta^{15}N$ values expressed in cultures that use the Mo-nitrogenase Nif; light blue, extension of this range seen in some wild types and thermophilic organisms using Nif; green, range observed in cultures using pure Vnf or Anf[7,17,18]. Almost all data points plot within the range of Mo-based nitrogen fixation. Outliers in fluvial sediments may reflect local redox chemistry.
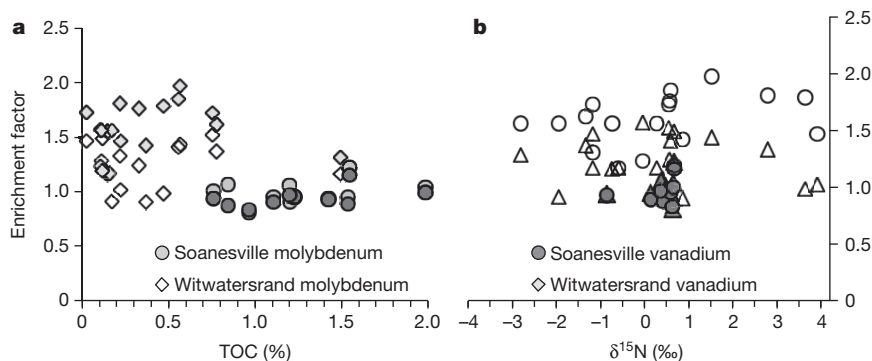
Figure 3 | TOC and δ15N versus Mo and V enrichment factors. a, TOC; **b**, δ15N. Enrichment factors are calculated relative to average upper continental crust (Methods). All values are relatively close to 1.0, indicating insignificant

authigenic enrichments of V or Mo in sediments. No trends are observed, indicating that availability of Mo and V did not influence either TOC or δ15N.

larger fractionations[22]. NO_x species produced by lightning were close to 0‰ in one laboratory experiment[23], but theoretical models and measurements of NO_x in the atmosphere indicate values between −5‰ and −15‰ (ref. 24), making this mechanism incompatible with our data. A significant hydrothermal NH_4^+ flux can also be ruled out, because it would not have been accessible in fluvial settings such as those of the Fortescue and argillaceous Witwatersrand rocks, and the similarity in δ15N$_{bulk}$ values between fluvial and marine samples suggests that nitrogen was supplied by the same mechanism. Another possible explanation for the δ15N$_{bulk}$ values could be nitrogen fixation by Vnf and/or Anf to −6‰, followed by a scenario similar to the modern ocean, namely quantitative nitrification, then partial denitrification in the water column imparting a positive fractionation to residual nitrate, and finally nearly quantitative assimilation of that nitrate into biomass. Today, this process imparts a positive fractionation to the resulting biomass of 5–7‰ compared with the initial fixed nitrogen. This could conceivably result in δ15N$_{bulk}$ values around 0‰ if Vnf and/or Anf were the dominant nitrogenases. In that case, however, our data should show a larger spread, as seen in the modern ocean (−2‰ to +15‰) where the extent of N_2 fixation and denitrification varies in response to climatic zonation[25] and ocean circulation[26]. Thus, given the small variance around 0‰ ($\sigma = 1.2‰$) shown by samples from different environments over a long time interval (450 Myr), this explanation seems unlikely. In particular, a standard deviation of only 0.5‰ among the Soanesville Group samples suggests a single simple fractionation mechanism.

Overall, our data point to an ecosystem where the supply of nitrogen was limited but sufficient to sustain productive microbial ecosystems. The most plausible explanation for nitrogen acquisition is therefore biological N_2 fixation. Mo was probably the dominant catalyst, indicated by the small variance of δ15N$_{bulk}$ values around 0‰ that is especially noticeable in the Soanesville Group; scatter may be due to varying Fe availability[17]. The alternative nitrogenases Vnf and Anf, and probably also potential precursor enzymes with low catalytic efficiency, should have caused larger or more variable fractionations than those observed. Modern Mo-dependent diazotrophs are capable of fixing N_2 down to approximately 1 nM [Mo], or about 1% of modern marine concentrations[27]. These low levels may have been supplied to mid-Archaean rivers and oceans by hydrothermal systems or localized oxidative weathering[21,28,29]. Mo uptake may have been further enhanced by molybdophores[30], if they had already evolved. Thus, the absence of appreciable Mo enrichments in these sediments (Fig. 3) does not preclude Mo bioavailability (Supplementary Information 3.2).

Our results have several important implications. First, the nitrogen cycle was predominantly anaerobic before 2.7 Gyr. Any biogenic O_2 (ref. 29) must therefore have had a short lifetime, and oxidative weathering[28] was probably restricted to environments where O_2 was locally produced. Second, nitrogenase is evidently a more ancient enzyme than recent models suggest[5]. Our data from the Soanesville Group constrain

the minimum age to about 3.2 Gyr and indicate that Mo was the most widely used cofactor. If so, Mo must have been sufficiently bioavailable in the mid-Archaean to serve as an enzymatic catalyst. This supports phylogenetic evidence for the early evolution of other Mo-containing enzymes[12]. Discrepancies with recent studies suggesting a Proterozoic origin of nitrogenase[5] may be due to either miscalibration of molecular clocks or lateral gene transfer that was unaccounted for in phylogenetic analyses. Lastly, the early evolution of Mo-nitrogenase and efficient nitrogen fixation would have averted the 'nitrogen crisis' predicted to have occurred in the Archaean or early Palaeoproterozoic[3,4], thus allowing biological evolution and proliferation to proceed apace.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Brandes, J. A. *et al.* Abiotic nitrogen reduction on the early Earth. *Nature* **395,** 365–367 (1998).
2. Tian, F., Kasting, J. F. & Zahnle, K. Revisiting HCN formation in Earth's early atmosphere. *Earth Planet. Sci. Lett.* **308,** 417–423 (2011).
3. Navarro-Gonzalez, R., McKay, C. P. & Nna Mvondo, D. A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature* **412,** 61–64 (2001).
4. Kasting, J. F. & Siefert, J. L. Biogeochemistry: the nitrogen fix. *Nature* **412,** 26–27 (2001).
5. Boyd, E. S. *et al.* A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology* **9,** 221–232 (2011).
6. Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The natural history of nitrogen fixation. *Mol. Biol. Evol.* **21,** 541–554 (2004).
7. Zhang, X., Sigman, D. M., Morel, F. M. & Kraepiel, A. M. Nitrogen isotope fractionation by alternative nitrogenases and past ocean anoxia. *Proc. Natl Acad. Sci. USA* **111,** 4782–4787 (2014).
8. Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C. & Dixon, R. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13,** 162 (2012).
9. Boyd, E. S., Hamilton, T. L. & Peters, J. W. An alternative path for the evolution of biological nitrogen fixation. *Front. Microbiol.* **2,** http://dx.doi.org/10.3389/fmicb.2011.00205 (2011).
10. Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297,** 1137–1142 (2002).
11. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452,** 456–459 (2008).
12. Schoepp-Cothenet, B. *et al.* The ineluctable requirement for the trans-iron elements molybdenum and/or tungsten in the origin of life. *Sci. Rep.* **2,** 1–5 (2012).
13. Garvin, J., Buick, R., Anbar, A. D., Arnold, G. L. & Kaufman, A. J. Isotopic evidence for an aerobic nitrogen cycle in the latest Archean. *Science* **323,** 1045–1048 (2009).
14. Godfrey, L. V. & Falkowski, P. G. The cycling and redox state of nitrogen in the Archaean ocean. *Nature Geosci.* **2,** 725–729 (2009).
15. Thomazo, C. & Papineau, D. Biogeochemical cycling of nitrogen on the early Earth. *Elements* **9,** 345–351 (2013).
16. Quan, T. M. & Falkowski, P. G. Redox control of N:P ratios in aquatic ecosystems. *Geobiology* **7,** 124–139 (2009).
17. Zerkle, A., Junium, C. K., Canfield, D. E. & House, C. H. Production of 15N-depleted biomass during cyanobacterial N_2-fixation at high Fe concentrations. *J. Geophys. Res.* **113,** http://dx.doi.org/10.1029/2007JG000651 (2008).

18. Nishizawa, M., Miyazaki, J., Makabe, A., Koba, K. & Takai, K. Physiological and isotopic characteristics of nitrogen fixation by hyperthermophilic methanogens: key insights into nitrogen anabolism of the microbial communities in Archean hydrothermal systems. *Geochim. Cosmochim. Acta* **138,** 117–135 (2014).
19. Marty, B., Zimmermann, L., Pujol, M., Burgess, R. & Philippot, P. Nitrogen isotopic composition and density of the Archean atmosphere. *Science* **342,** 101–104 (2013).
20. Stüeken, E. E. A test of the nitrogen-limitation hypothesis for retarded eukaryote radiation: nitrogen isotopes across a Mesoproterozoic basinal profile. *Geochim. Cosmochim. Acta* **120,** 121–139 (2013).
21. Guy, B. M. *et al.* A multiple sulfur and organic carbon isotope record from non-conglomerate sedimentary rocks of the Mesoarchean Witwatersrand Supergroup, South Africa. *Precambr. Res.* **216–219,** 208–231 (2012).
22. Kuga, M. *et al.* Nitrogen isotopic fractionation during abiotic synthesis of organic solid particles. *Earth Planet. Sci. Lett.* **393,** 2–13 (2014).
23. Hoering, T. The isotopic composition of the ammonia and the nitrate ion in rain. *Geochim. Cosmochim. Acta* **12,** 97–102 (1957).
24. Moore, H. The isotopic composition of ammonia, nitrogen dioxide and nitrate in the atmosphere. *Atmos. Environ.* **11,** 1239–1243 (1977).
25. Galbraith, E., Kienast, M. & NICOPP. The acceleration of oceanic denitrification during deglacial warming. *Nature Geosci.* **6,** 579–584 (2013).
26. Straub, M. *et al.* Changes in North Atlantic nitrogen fixation controlled by ocean circulation. *Nature* **501,** 200–203 (2013).
27. Glass, J. B., Wolfe-Simon, F., Elser, J. J. & Anbar, A. D. Molybdenum-nitrogen co-limitation in freshwater and coastal heterocystous cyanobacteria. *Limnol. Oceanogr.* **55,** 667–676 (2010).
28. Crowe, S. A. *et al.* Atmospheric oxygenation three billion years ago. *Nature* **501,** 535–538 (2013).
29. Planavsky, N. J. *et al.* Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nature Geosci.* **7,** 283–286 (2014).
30. Liermann, L. J., Guynn, R. L., Anbar, A. & Brantley, S. L. Production of a molybdophore during metal-targeted dissolution of silicates by soil bacteria. *Chem. Geol.* **220,** 285–302 (2005).

Author Contributions R.B. collected samples from the Soanesville Group; B.M.G. collected samples from the Witwatersrand Supergroup; R.B. and E.E.S. collected samples from the Fortescue Group and Nullagine Group; E.E.S. and M.C.K. performed the analyses; E.E.S. drafted the manuscript with advice and contributions from R.B.; B.M.G. and M.C.K. commented on the draft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.E.S. (evast@uw.edu).

METHODS

No statistical methods were used to predetermine sample size.

Sample preparation. Sample preparation followed established protocols[20]. First, all outer surfaces were cut or polished off, then the rocks were crushed into sub-centimetre chips and sonicated sequentially in hexane, methanol, and deionized water (18 MΩ) for a few minutes each to remove any modern contaminants. The chips were dried at 60 °C in an oven overnight and then pulverized in an aluminium oxide ceramic puck mill. The mill was cleaned with pre-combusted sand, deionized water, and methanol in between samples. Rock powders were de-carbonated with 6 N HCl for 3 days at 60 °C, where the acid was refreshed after 1 day and stirred after 2 days. The de-carbonated samples were washed with deionized water three times to remove residual acid and dried at 60 °C. All glassware used during sample preparation and storage was combusted in a muffle-furnace at 500 °C overnight.

Kerogen isolates were prepared following the HF–BF_3 maceration protocol of ref. 31. Five to six grams of untreated rock powder were weighed into a 250 ml Teflon bottle, mixed with 100 ml deionized water (18 MΩ) and 100 ml concentrated HF (reagent grade), and placed into a heated shaking water bath at 55 °C overnight. The next day, the HF was decanted after centrifugation for 45 min. Then a mixture of 62.5 g H_3BO_3 (reagent grade), 100 ml H_2O, and 100 ml HF was added to each sample and placed back into the heated shaking water bath overnight. This mixture results in the formation of BF_3, which efficiently removes fluoride precipitates[31]. The BF_3 was decanted the following day after centrifugation, the sample was washed three times with deionized water, and then transferred with about 10 ml of deionized water into a scintillation vial. The vials were then placed into a freezer overnight and desiccated with a freeze-drier over the following day. Freezing the liquid first minimizes loss of kerogen by splashing during evacuation in the freeze-drier. Reproducibility of kerogen extraction was tested with an in-house rock standard from the 2.5 Gyr Mount McRae Shale (UW-McRae) and was around 0.1‰ for $\delta^{15}N_{ker}$. We were not able to extract kerogen from all of our samples owing to limitations in the amount of sample available and their sometimes low TOC content.

Isotopic analyses. All analyses were done with an elemental analyser (Costech ECS 4010) coupled to a continuous-flow isotope-ratio mass spectrometer (Finnigan MAT253) via a ThermoFinnigan Conflo III[13,20,32]. Powders were weighed into 9 mm × 5 mm tin capsules (Costech) and flash-combusted with 10 ml O_2 ('macro' setting) at 1 bar and 1,000 °C. Resulting N_2 and CO_2 gases were purified with Cr_2O_3 (conversion of trace CO to CO_2) at 1,000 °C, silvered Co_3O_4 (capture of sulphur gases and halogens) at 1,000 °C, and Cu (conversion of trace NO_x to N_2) at 650 °C. Traces of H_2O were scrubbed with $Mg(ClO_4)_2$ at room temperature (20–25 °C). N_2 and CO_2 were then separated chromatographically at 40 °C and analysed sequentially. For analyses of kerogen isolates, CO_2 gas was scrubbed from the samples with soda lime at room temperature, which allowed combustion of larger amounts of kerogen without any risk of saturating the detector cups in the mass spectrometer. Kerogen was mixed with V_2O_5 in a 1:1 ratio by mass to enhance combustion efficiency, although subsequent tests showed that samples also combusted well without an additional combustion aid (discussed below; Extended Data Fig. 3).

All measurements were corrected for instrumental mass bias with two previously calibrated in-house standards (glutamic acid GA-2 and dried salmon SA, calibrated against USGS41 and USGS40 (ref. 33)), which were analysed after each set of five samples. Accuracy and long-term reproducibility were monitored with a second glutamic acid standard (GA-1) and the UW-McRae rock standard. In all runs, accuracy for $\delta^{15}N$ was better than 0.3‰ (average absolute value 0.1‰) and average long-term precision over 12 months was also 0.3‰. For $\delta^{13}C_{org}$, the average long-term accuracy and average precision were both 0.1‰. Samples were analysed in two or three replicates of the same decarbonated powder aliquot; standard deviations are provided in Supplementary Table A2. Results are expressed in delta notation ($\delta^{15}N = \{[^{15}N/^{14}N]_{sample}/[^{15}N/^{14}N]_{standard} - 1\}$; $\delta^{13}C_{org} = \{[^{13}C/^{12}C]_{sample}/[^{13}C/^{12}C]_{standard} - 1\}$); reference standards were atmospheric N_2 for nitrogen isotopes and Vienna PeeDee Belemnite for carbon isotopes. N and C concentrations were reproducible with average precision of 1.7% and 1.2%, respectively. All whole-rock samples and kerogen isolates were analysed in duplicates or triplicates.

Combustion efficiency test. It has been noted that flash combustion may be inefficient in oxidizing organic carbon quantitatively to CO_2 such that any residual CO may interfere with N_2 in the mass spectrometer and increase the measured ratio of masses 29 to 28, leading to artificially higher $\delta^{15}N$ values[34,35]. Furthermore, NH_4^+ bound to phyllosilicate minerals may not be quantitatively released from the sample[36]. In both cases V_2O_5 has been proposed as a combustion aid; however, it can contain traces of nitrogen itself that need to be subtracted.

In a previous study[20] using the same analytical protocol and facilities, we demonstrated that we could achieve good reproducibility in $\delta^{15}N$ ($1\sigma = 0.11$‰) over a range of V_2O_5/sample ratios from 0 to 1.7 for a shale sample from the 1.5 Gyr Belt Supergroup. In addition, the measured total nitrogen concentrations showed no correlation with the V_2O_5/sample ratio ($r^2 = 0.008$) and had a high precision (relative error 0.96%). Kerogen extraction and mass balance showed that 25% of all nitrogen in this sample was organic-bound (with $\delta^{15}N = +4.7 \pm 0.3$‰) and 75% was silicate-bound (with $\delta^{15}N = +5.1 \pm 0.4$‰). The good reproducibility of both isotopic ratio and total nitrogen abundance over a range of sample quantities from 50 mg to 150 mg implies that even in the absence of V_2O_5 the flash combustion technique quantitatively oxidizes all organic- and silicate-bound nitrogen in this type of sample, where the majority of nitrogen is silicate-bound. Furthermore, the data suggest that CO interferences are insignificant for up to 1.66 mg organic carbon, which was the highest amount combusted with this sample (TOC = 1.1%, molar C/N_{bulk} ratio = 29).

In the course of this study, we conducted a second test with the UW-McRae standard (TOC = 7.7%, TN = 0.11%) to verify that this conclusion holds at high organic carbon abundances. We analysed aliquots of powder ranging from 1.6 mg to 61.2 mg, equivalent to 0.1–4.7 mg organic carbon, and achieved good reproducibility for both $\delta^{15}N$ ($1\sigma = 0.19$‰) and total nitrogen (relative error 5.8%), without addition of V_2O_5 and at constant O_2 supply during flash combustion. If traces of CO were produced, it would be most significant at large sample sizes—that is, at small O_2/sample ratios. However, the absence of any significant trend in the data (Extended Data Fig. 3) indicates that, up to at least 4.7 mg organic carbon, interferences with CO can be disregarded.

Elemental analyses. We submitted aliquots of powder to ALS for major, minor, and trace element analyses (ALS method ME-MS61). Samples were digested with HF, HNO_3, and $HClO_4$, then analysed in dilute HCl with inductively coupled plasma–atomic emission spectrometry and inductively coupled plasma–mass spectrometry. We calculated enrichment factors (EF) for Mo and V (EF = $([X]/[Al])_{sample}/([X]/[Al])_{reference}$, where [X] is the concentration of either Mo or V) with reference to average upper continental crust[37] and average Archaean crust[38] (Fig. 3 and Extended Data Fig. 2).

31. Robl, T. L. & Davis, B. H. Comparison of the HF-HCl and HF-BF₃ maceration techniques and the chemistry of resultant organic concentrates. *Org. Geochem.* **20,** 249–255 (1993).

32. Schoepfer, S. D., Henderson, C. M., Garrison, G. H. & Ward, P. D. Cessation of a productive coastal upwelling system in the Panthalassic Ocean at the Permian–Triassic boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **313-314,** 181–188 (2012).

33. Qi, H., Coplen, T. B., Geilmann, H., Brand, W. A. & Böhlke, J. K. Two new organic reference materials for δ¹³C and δ¹⁵N measurements and a new value for the δ¹³C of NBS 22 oil. *Rapid Commun. Mass Spectrom.* **17,** 2483–2487 (2003).

34. Beaumont, V., Agrinier, P., Javoy, M. & Robert, F. Determination of the CO contribution to the ¹⁵N/¹⁴N ratio measured by mass spectrometry. *Anal. Chem.* **66,** 2187–2189 (1994).

35. Papineau, D. *et al.* High primary productivity and nitrogen cycling after the Paleoproterozoic phosphogenic event in the Aravalli Supergroup, India. *Precambr. Res.* **171,** 37–56 (2009).

36. Bräuer, K. & Hahne, K. Methodical aspects of the ¹⁵N-analysis of Precambrian and Palaeozoic sediments rich in organic matter. *Chem. Geol.* **218,** 361–368 (2005).

37. Rudnick, R. L. & Gao, S. Composition of the continental crust. *Treat. Geochem.* **4,** 1–51 (2014).

38. Taylor, S. R. & McLennan, S. M. in *Encyclopedia of Physical Science and Technology* 3rd edn (ed. Meyers, R. A.) 697–719 (Academic, 2001).
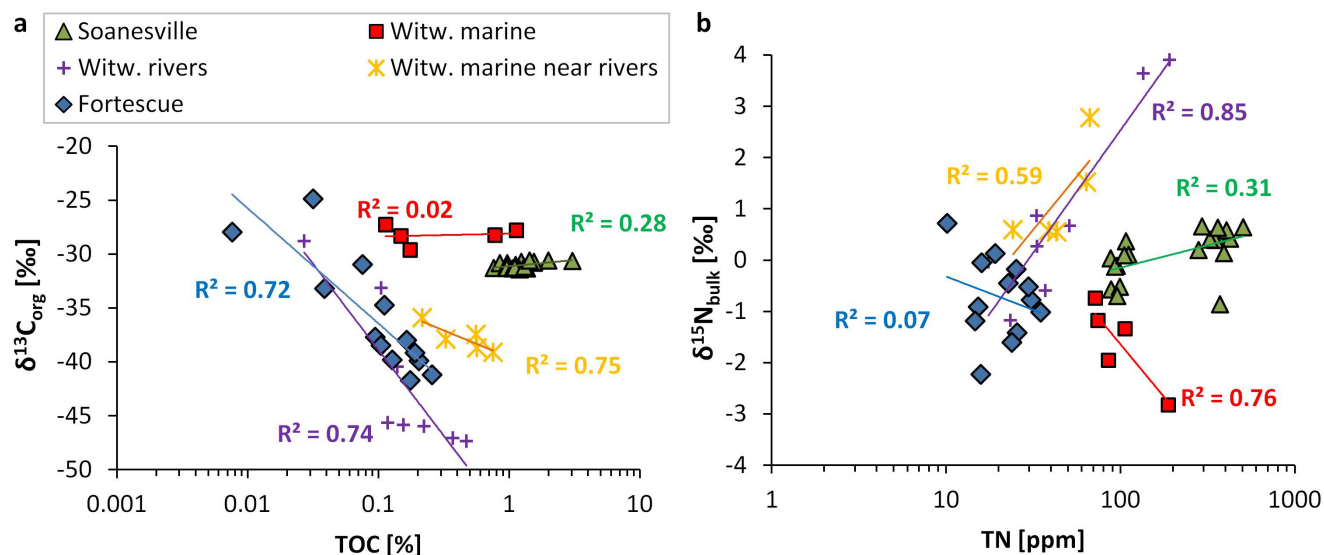
Extended Data Figure 1 | Organic carbon (a) and bulk nitrogen (b) isotopes as a function of total abundances. Correlation coefficients refer to data points and trend lines of the same colour. See Supplementary Information 3 for discussion. Each sample was analysed in two or three replicates.
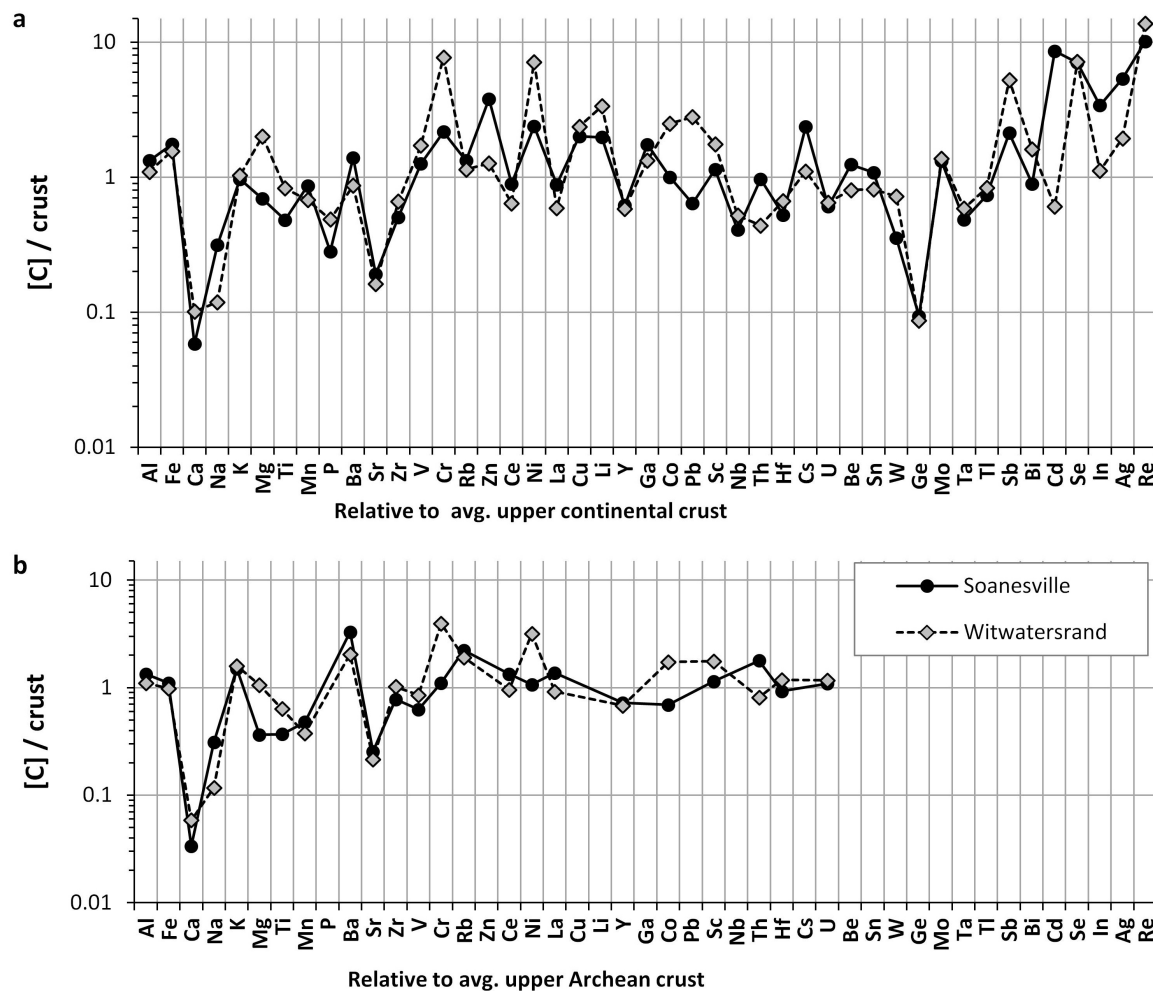
Extended Data Figure 2 | Crust-normalized elemental abundances.
a, Normalized relative to average upper continental crust; **b**, normalized relative to average upper Archaean crust (Methods). Some data points in **b** are missing because concentrations for those elements in average upper Archaean crust were not provided. In both parts, elements are sorted by relative abundance in upper continental crust.
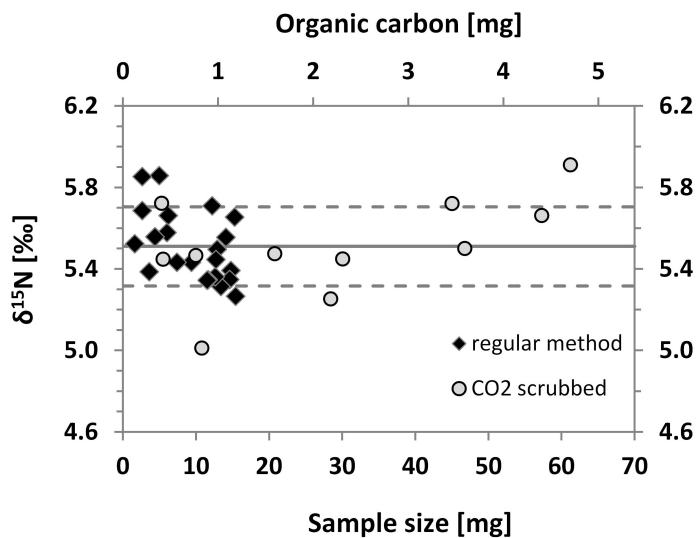
Extended Data Figure 3 | Reproducibility of the UW-McRae in-house standard over a range of sample sizes. Filled diamonds, analysed with regular CN method; grey circles, analysed with a soda-lime column to scrub CO_2 from the gas stream; horizontal lines, total average (solid) ± 1 s.d. (dashed).

Extended Data Table 1 | Comparison of kerogen and whole-rock data for samples with sufficient material from the Soanesville Group and the Witwatersrand Supergroup

Sample	$\delta^{15}N_{bulk}$ [‰]	TN [ppm]	TOC [%]	C/N_{bulk} [atom.]	$\delta^{15}N_{ker}$ [‰]	σ [‰]	C/N_{ker} [atom.]	$f(N_{ker})$ [%]	$f(N_{sil})$ [%]
Soanesville Group:									
96049	0.48	359	1.20	39	-1.29	0.10	274	14	86
96050	0.40	348	0.76	25	-0.99	0.07	252	10	90
96051	0.42	423	1.11	31	-1.63	0.10	278	11	89
96052	0.40	325	0.85	30	-1.57	0.01	266	11	89
96053	0.58	405	1.23	35	-1.47	0.31	267	13	87
96054	0.14	392	1.54	46	-2.08	0.02	282	16	84
96055	0.64	503	1.98	46	-1.75	0.01	289	16	84
96056	0.63	360	0.96	31	-2.12	0.16	287	11	89
97SSD32-06	0.66	292	1.55	62	-1.40	0.01	297	21	79
97SSD32-21	0.37	107	1.20	132	-2.24	0.23	302	43	57
97SSD51-03	-0.86	371	1.42	45	-2.25	0.06	267	17	83
Witwatersrand Supergroup:									
BABI_1119	-0.59	37	0.16	49	-2.40	0.01	673	7	93
BABI_115	3.64	135	0.47	41	-0.48	0.30	394	10	90
TW6_1749.9	0.57	39	0.33	102	-2.17	0.72	1279	8	92
TW6_1750.2	1.53	64	0.56	104	-1.79	0.05	1251	8	92
TF1_4400	-1.17	74	1.13	171	-2.28	0.64	1910	9	91
TF1_4440	-1.33	106	0.78	86	-4.44	0.28	1781	5	95

Whole-rock data are repeated from Supplementary Table A2.

LETTER

An epigenome–wide association study of total serum immunoglobulin E concentration

Liming Liang[1], Saffron A. G. Willis-Owen[2]*, Catherine Laprise[3]*, Kenny C. C. Wong[2], Gwyneth A. Davies[4], Thomas J. Hudson[5,6], Aristea Binia[2], Julian M. Hopkin[4], Ivana V. Yang[7], Elin Grundberg[8], Stephan Busche[8], Marie Hudson[9], Lars Rönnblom[10], Tomi M. Pastinen[8,11], David A. Schwartz[7], G. Mark Lathrop[8]§, Miriam F. Moffatt[2]§ & William O. C. M. Cookson[2]§

Immunoglobulin E (IgE) is a central mediator of allergic (atopic) inflammation. Therapies directed against IgE can alleviate hay fever[1] and allergic asthma[1,2]. Genetic association studies have not yet identified novel therapeutic targets or pathways underlying IgE regulation[3–6]. We therefore surveyed epigenetic associations between serum IgE concentrations and methylation at loci concentrated in CpG islands genome wide in 95 nuclear pedigrees, using DNA from peripheral blood leukocytes. We validated positive results in additional families and in subjects from the general population. Here we show replicated associations—with a meta-analysis false discovery rate less than 10^{-4}—between IgE and low methylation at 36 loci. Genes annotated to these loci encode known eosinophil products, and also implicate phospholipid inflammatory mediators, specific transcription factors and mitochondrial proteins. We confirmed that methylation at these loci differed significantly in isolated eosinophils from subjects with and without asthma and high IgE levels. The top three loci accounted for 13% of IgE variation in the primary subject panel, explaining the tenfold higher variance found compared with that derived from large single-nucleotide polymorphism genome-wide association studies[3,4]. This study identifies novel therapeutic targets and biomarkers for patient stratification for allergic diseases.

Asthma, atopic dermatitis (eczema) and hay fever are IgE-related diseases that are increasing in prevalence and are a major source of disability. Systematic knowledge of IgE production is limited, beyond the regulation of IgE creation in B cells by interleukin (IL)-4 released from T_H2 cells and eosinophils[7]. Genome-wide association studies show polymorphisms in *STAT6*, *FCER1A*, the *IL4/RAD50* locus and the major histocompatibility complex (MHC) to be associated with high IgE concentrations[3–6], but these single-nucleotide polymorphisms (SNPs) combined account for only 1–2% of the variation in serum IgE[4].

CpG methylation is associated with gene silencing and the patterning of gene expression that determines cellular types and functions[8],

and islands of CpG (CGIs) sequences are positioned near the promoters of 40% of human genes[9]. *IL4* expression has been related to upstream epigenetic variation in DNA methylation in T cells[10], encouraging us to search genome wide for other CGIs associated with IgE serum concentrations.

We used Illumina HumanMethylation27 arrays to target individual CpG sites (loci) within proximal promoter regions of 14,475 genes. The panel is enriched for genomic regions regulating expression, but does not cover all functionally important CpG sites. We excluded from downstream analyses any loci with SNPs overlapping the Illumina probe sequence, and established that direct bisulphite pyrosequencing correlates robustly with the array used in this study (Extended Data Fig. 1) and elsewhere[11].

We investigated nuclear families from the Medical Research Council Asthma (MRCA) panel, in which we have previously carried out genome-wide SNP association studies for IgE levels and asthma[12]. The panel contained 355 subjects (183 male) with a mean age in children of 12.2 years (ranging from 2 to 39 years) and in adults of 42 years (27 to 61 years) (Table 1). One-hundred and thirteen children had doctor-diagnosed asthma. We found replication cohorts in 149 Caucasian subjects selected equally from the top and bottom deciles of IgE distribution in 1,614 unselected volunteers for the Poblogaeth Asthma Prifysgol Abertawe (PAPA) study (students and staff from Swansea University)[13]; and in 160 subjects in an asthmatic family panel from the Saguenay–Lac-Saint-Jean region (SLSJ) of Québec[14], with a mean age in children of 16 years (ranging from 5 to 50 years; 40 with doctor-diagnosed asthma) and adults of 44 years (31 to 79 years) (Table 1).

We fitted models with $\log_e$-normalized IgE (ln(IgE)) as a dependent variable and methylation status for each Illumina probe as a predictor, with age, sex, parental status, interactions and batch identifiers as covariates. We identified 34 loci with a false discovery rate (FDR) < 0.01 (Fig. 1 and Supplementary Table 1) in 32 different CGIs in the MRCA

Table 1 | Subject characteristics

	MRCA (discovery)	PAPA (1st replication)	SLSJ (2nd replication)
Number	355	149	160
Age (mean (range))	28 (2–61)	21 (18–30)	29 (5–79)
N (%) female	17 (48.5%)	72 (48.3%)	80 (50.0%)
N (%) asthmatic	175 (49.3%)	34 (22.8%)	69 (43.1%)
N (%) smoker	45 (12.7%)	33 (22.1%)	28 (17.5%)
Eosinophil count (mean ± s.e.m.) per μl	406 ± 383	246 ± 214	242 ± 205
Geometric mean serum IgE (range) (IU l^{-1})†	320 (1–4,999)	663 (0–18,800)	412 (2–7,653)

s.e.m., standard error of the mean.
* Normal range, <350 cells per μl.
† Normal range, <100 IU l^{-1}

[1]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [2]National Heart and Lung Institute, Imperial College, London SW3 6LY, UK. [3]Université du Québec à Chicoutimi, Saguenay, Québec G7H 2B1, Canada. [4]Institute of Life Science, College of Medicine, Swansea University, Swansea SA2 8PP, UK. [5]Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. [6]Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Ontario M5S 1A1, Canada. [7]University of Colorado School of Medicine and National Jewish Health, Denver, Colorado 80206, USA. [8]Department of Human Genetics, McGill University and Génome Québec Innovation Centre, Montréal H3A 1B1, Canada. [9]Jewish General Hospital and Lady Davis Research Institute, Montréal H3T 1E2, Canada. [10]Department of Medical Sciences, SciLifeLab, Uppsala University, Uppsala SE-751 44, Sweden. [11]Department of Medical Genetics, McGill University Health Centre, Montréal H3H 1P3, Canada.
*These authors contributed equally to this work.
§These authors jointly supervised this work.
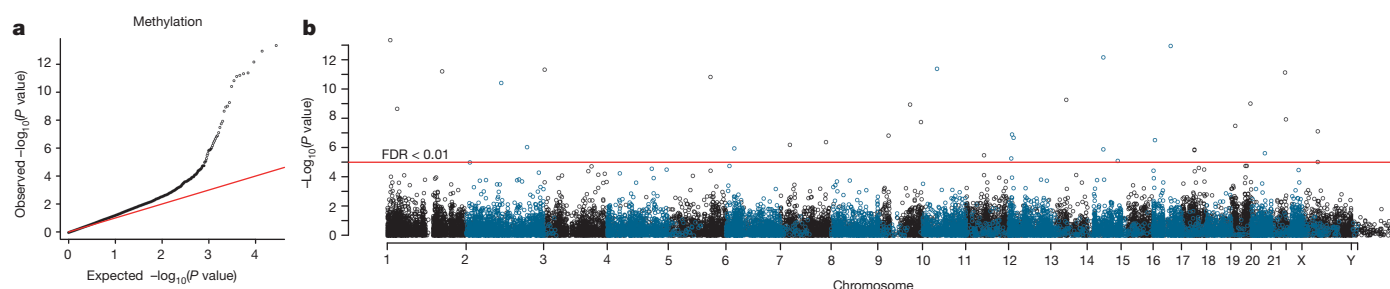
Figure 1 | Manhattan plot of the results of the genome-wide methylation association study. a, b, The results of genome-wide association testing to CGIs are shown for 27,000 loci in 355 subjects from the MRCA panel of families. **a,** The QQ plot showing observed versus expected $-\log_{10}(P\text{ values})$ for association at all loci. **b,** Manhattan plot showing chromosomal locations of $-\log_{10}(P\text{ values})$ for association at each locus. The red line illustrates the threshold for an FDR < 0.01.

panel. After replication in the PAPA and SLSJ panels, a meta-analysis combining the results identified 36 loci with FDR $< 10^{-4}$ and 62 loci with an FDR < 0.005 (Table 2 and Supplementary Table 1). All loci showed associations with the same anti-correlated direction in the three data sets (Table 2). A radioallergosorbent test (RAST) index, quantifying IgE against common allergens[15], showed similar but non-independent associations, suggesting common regulation of total and specific IgE. Testing of models with asthma as the dependent variable showed only *LPCAT2* (also known as *AYTL1*) and *ZNF22* to be associated with asthma independently of IgE levels ($P = 7.7 \times 10^{-5}$ and $P = 1.8 \times 10^{-4}$).

The variable methylation site upstream of *IL4* has a well-studied effect on IL-4 production[16,17] and IgE regulation, with methylation anti-correlated with expression in the same direction as in our study. We looked for SNP associations at this locus by imputation with the 1000

Genomes project phase 1 SNPs and indels in all three panels, analysing the 20,746 variants within 1 megabase upstream or downstream of the *IL4* 5′ untranslated region. We found no significant SNP associations with IgE after accounting for multiple testing.

We carried out Mendelian randomization to test for a causal effect of IL-4 methylation on IgE[18], choosing the SNP showing the strongest association to methylation at the IL-4 CpG probe (cg26787239) as the instrumental variable. The first stage *F*-test statistics for the MRCA and SLSJ panels ($F = 16.4$ and 26.2) indicated effects strong enough to ensure the validity of the method. In the MRCA panel, the association between the instrument SNP (*rs12311504*) and IgE before adjusting for IL-4 methylation was $P = 0.03$, and $P = 0.53$ after adjustment, indicating that methylation mediated most of the SNP effect. The meta-analysis *P* for a causal effect was 6.8×10^{-4}, suggesting that the locus

Table 2 | Meta-analysis of association of total serum IgE concentration in three subject panels

Probe	Symbol	Function	MRCA (P value)	SLSJ (P value)	PAPA (P value)	Meta-analysis (P value)
cg01998785	LPCAT2	Lysophospholipid metabolism	1.2×10^{-13}	8×10^{-3}	9.6×10^{-6}	1.2×10^{-18}
cg10159529	IL5RA	Cytokine signalling	5.1×10^{-12}	2.1×10^{-4}	7.2×10^{-5}	2.2×10^{-18}
cg01614759	ZNF22	Transcription factor	4.4×10^{-12}	3.4×10^{-3}	2.8×10^{-6}	2.8×10^{-18}
cg15996947	L2HGDH	Mitochondrial oxidoreductase	7.4×10^{-13}	1.0×10^{-2}	3.7×10^{-5}	2.8×10^{-17}
cg26787239	IL4	Cytokine signalling	1.6×10^{-11}	1.3×10^{-3}	4.8×10^{-4}	3.2×10^{-16}
cg18783781	SLC25A33	Mitochondrial transport: dendritic-cell endocytosis	5.0×10^{-14}	3.4×10^{-2}	6.2×10^{-3}	4.4×10^{-15}
cg13221796	RB1	Transcription factor	5.7×10^{-10}	6.0×10^{-2}	4.9×10^{-6}	2.5×10^{-14}
cg01770400	SERPINC1	Anti-thrombin	6.6×10^{-12}	7.0×10^{-3}	1.3×10^{-2}	5.3×10^{-14}
cg02643667	TFF1	Mucus-stabilizing secreted protein	7.9×10^{-12}	1.3×10^{-1}	2.6×10^{-3}	7.6×10^{-13}
cg21627181	SLC17A4	Sodium/phosphate co-transporter	1.1×10^{-6}	1.4×10^{-4}	3.8×10^{-4}	1.1×10^{-12}
cg20189937	L2HGDH	Mitochondrial oxidoreductase	1.3×10^{-6}	2.4×10^{-3}	3.2×10^{-5}	2.6×10^{-12}
cg26457013	TMEM86B	Lysoplasmalogenase: phospholipid metabolism	1.0×10^{-9}	9.5×10^{-2}	9.0×10^{-5}	7.1×10^{-12}
cg20503329	COL15A1	Cell shape, motility, adhesion	1.2×10^{-9}	3.0×10^{-1}	8.8×10^{-5}	9.5×10^{-12}
cg03693099	CEL	Secreted carboxyl ester lipase	1.8×10^{-8}	1.1×10^{-1}	8.3×10^{-5}	1.3×10^{-11}
cg00079056	SPINK4	Serine peptidase inhibitor	1.5×10^{-7}	6.4×10^{-2}	3.1×10^{-5}	1.8×10^{-11}
cg09676390	ADARB1	Pre-mRNA editing of the glutamate receptor	1.2×10^{-8}	7.0×10^{-2}	8.0×10^{-4}	3.1×10^{-11}
cg15998761	MFSD6	MHC receptor homologue	9.3×10^{-7}	1.3×10^{-1}	1.7×10^{-4}	4.5×10^{-11}
cg25494227	TMEM52B	Transmembrane protein	1.3×10^{-7}	2.0×10^{-1}	1.2×10^{-5}	5.1×10^{-11}
cg11398517	FAM112A	Unknown	2.4×10^{-6}	7.4×10^{-3}	3.3×10^{-4}	1.0×10^{-10}
cg06690548	SLC7A11	Cystine/glutamate antiporter: dendritic-cell differentiation	2.7×10^{-5}	3.3×10^{-4}	1.1×10^{-3}	1.8×10^{-10}
cg17784922	KEL	Metallo-endopeptidase	4.2×10^{-7}	7.9×10^{-3}	4.4×10^{-3}	2.1×10^{-10}
cg16050349	PIK3CB	Catalytic subunit for PI3K-β: activation of neutrophils	4.0×10^{-5}	1.7×10^{-3}	2.3×10^{-4}	3.2×10^{-10}
cg25636075	TMEM41A	Transmembrane protein	2.5×10^{-4}	5.1×10^{-5}	7.7×10^{-4}	3.9×10^{-10}
cg08404225	IL5RA	Cytokine signalling	2.3×10^{-4}	3.3×10^{-3}	8.4×10^{-6}	4.1×10^{-10}
cg09447105	PDE6H	Inhibitory subunit of cGMP phosphodiesterase	2.2×10^{-7}	1.3×10^{-7}	4.0×10^{-4}	5.3×10^{-10}
cg05215575	SEPT12	Cell shape, motility, adhesion	3.1×10^{-7}	2.1×10^{-1}	2.7×10^{-4}	1.2×10^{-9}
cg26136776	KLF1	Erythroid-specific transcription factor	3.3×10^{-8}	4.3×10^{-1}	6.6×10^{-4}	1.5×10^{-9}
cg17749520	ITGA2B	Platelet fibronectin receptor: role in coagulation	1.4×10^{-6}	2.5×10^{-2}	3.5×10^{-3}	1.8×10^{-9}
cg24459209	PRG3	Eosinophil major basic protein homologue	3.3×10^{-6}	2.9×10^{-2}	1.1×10^{-3}	1.8×10^{-9}
cg00002426	SLMAP	Sarcolemma-associated protein	7.9×10^{-5}	8.3×10^{-3}	1.6×10^{-4}	2.4×10^{-9}
cg15357945	PRG2	Eosinophil granule major basic protein	2.2×10^{-3}	2.8×10^{-5}	5.8×10^{-4}	3.1×10^{-9}
cg17582777	EFNA3	Receptor protein-tyrosine kinase	1.1×10^{-4}	3.1×10^{-1}	8.2×10^{-5}	8.6×10^{-9}
cg19881895	SLC43A3	Transmembrane protein	7.5×10^{-5}	2.8×10^{-3}	6.7×10^{-3}	1.6×10^{-8}
cg18254848	CLC	Lysophospholipid metabolism	1.8×10^{-5}	4.4×10^{-2}	4.6×10^{-3}	4.5×10^{-8}
cg21631409	ALDH3B2	Enzyme or kinase	2.3×10^{-4}	1.7×10^{-2}	1.2×10^{-3}	6.8×10^{-8}
cg00536175	GATA1	Eosinophil-specific transcription factor	7.9×10^{-8}	4.0×10^{-1}	5.1×10^{-2}	1.4×10^{-7}

Loci with an FDR for the meta-analysis $<10^{-4}$ are shown: a full list of significant associations can be found in Supplementary Table 1. MRCA contained 355 subjects, PAPA 149 and SLSJ 160. Markers are identified through their Illumina identification numbers and the associated gene symbol is derived from the Illumina annotation updated through PubMed. Note that two probes from *IL5RA* and from *L2HGDH* are associated to IgE concentrations.

represents a functionally validated epigenetic association with a complex phenotype.

Several loci were annotated to genes that encode proteins characteristic of eosinophils (Table 2 and Supplementary Table 1). *IL5RA* encodes a receptor that selectively stimulates eosinophil production and activation[19]; *CCR3* encodes the eosinophil eotaxin receptor; *IL1RL1* encodes the receptor for the eosinophil-activating cytokine IL-33; *PGR2* (also known as *GPR142*) encodes eosinophil granule major basic protein (*PRG2*); *PGR3* (also known as *GPR139*) is a *PGR2* homologue; and *GATA1* is an eosinophil transcription factor. We therefore tested whether methylation at our associated loci marked activation in eosinophils purified from peripheral blood, studying eight asthmatics with high serum IgE levels (>110 international units $(IU) l^{-1}$), eight asthmatics with low serum IgE ($<110 IU l^{-1}$) levels, and eight controls (mean age of all subjects 31 years (range 6–56 years), eight females and two current smokers). Asthmatics in both groups were on a maintenance regime of inhaled β-agonists, augmented with inhaled glucocorticoids during exacerbations.

We observed the lowest levels of methylation in the subjects with asthma and high IgE, and that methylation in asthmatics with low IgE was intermediate to controls (Fig. 2) ($P < 0.05$; Supplementary Table 1), supporting our initial results. Partitioning the data into high or low IgE groups led to similar conclusions. The range of variation for the principal loci was narrower in asthmatics with high IgE (Fig. 2) than in the other two groups, suggesting the enrichment of a distinctive eosinophil subset in atopic asthma. This is consistent with the recognized mixture of eosinophil populations in human blood and eosinophil activation in atopic disease[20]. Comparison of methylation levels in the MRCA panel

for our IgE-associated loci ($P < 0.001$) with an independent study of eosinophils isolated from normal subjects[21] confirmed correlations with methylation status ($R = 0.64$).

Lineage commitment to particular cell types is accompanied by specific methylation changes[8], and it has been suggested that DNA from mixtures of cells (such as peripheral blood leukocytes (PBLs)) will not support epigenome-wide association studies of complex diseases[21]. We extracted DNA from unfractionated populations of PBLs so that our methylation patterns reflect the numbers and the activity of different cells in each specimen. We further explored whether our associations to IgE reflected carriage in particular cell types by fitting regression models that included differential white cell counts. We identified partial associations with eosinophil numbers for all IgE-associated loci (Supplementary Table 2 and Fig. 3), consistent with independent effects on IgE from the numbers of eosinophils and the activity of the loci within eosinophils.

It is well recognized that the regulation of IgE production against particular antigens may reside in T cells and B cells as well as in eosinophils[22]. However, our regression models found that the top IgE associations were not accounted for by concomitant correlation with lymphocyte counts (Supplementary Table 2). We further examined the distribution of IgE-associated CpG loci in leukocyte subsets we isolated and in subsets from published data[21]. The results showed robustly in both data sets that low levels of methylation at the IgE-associated loci were confined to eosinophils (Extended Data Fig. 2).

Surrogate CpG markers that identify lymphocyte subsets can be used as an alternative to white cell counts in association models[23]. We also applied these methods to our data (Extended Data Table 1). This analysis
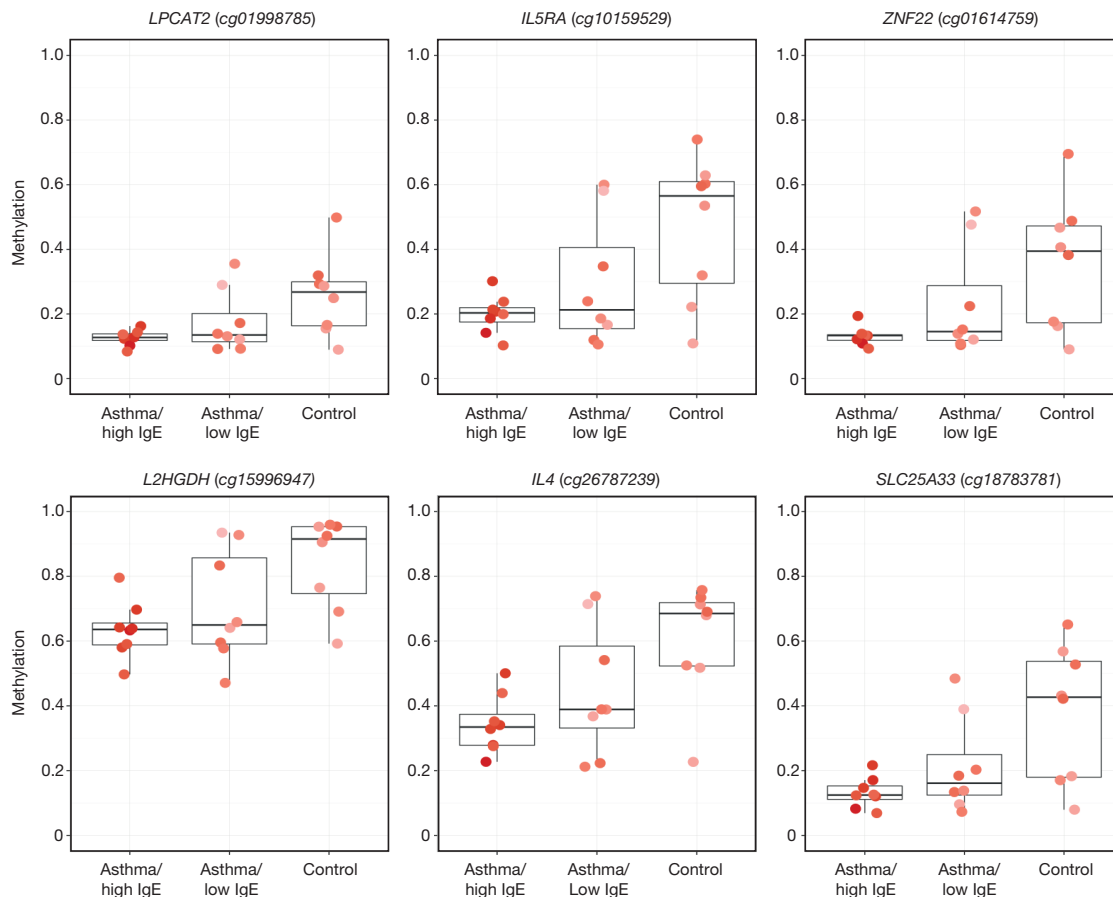
Figure 2 | Boxplots of methylation at selected CpG loci in isolated eosinophils from subjects with and without asthma and high total serum IgE concentrations ($>110 IU l^{-1}$). Subjects were recruited from the SLSJ population. Methylation (β) is shown on a scale of 0–1 for eight subjects in each group. Boxplots show means and interquartile ranges. The intensity of the data point colour is proportional to total serum IgE. All loci exhibited reduced variability and levels of methylation in the subjects with asthma and high IgE ($P < 0.05$).
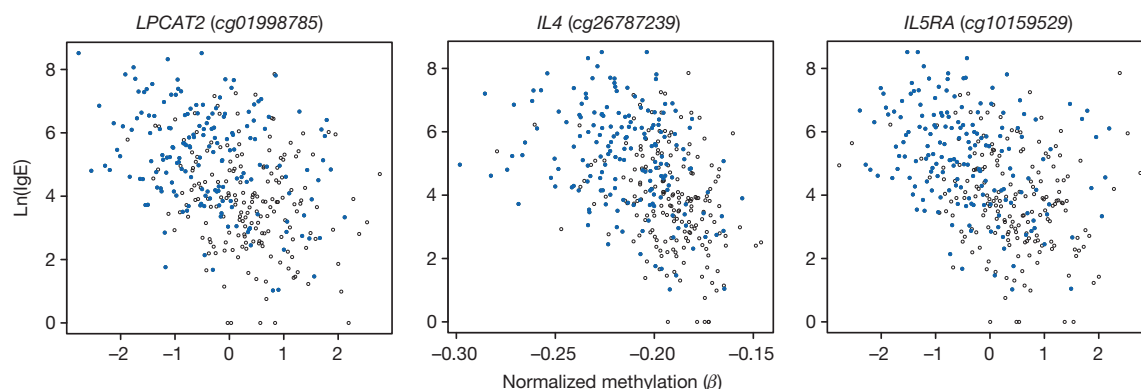
Figure 3 | Association of selected CpG loci to total serum IgE concentrations in the MRCA panel, partitioned by eosinophil counts. Methylation values for 355 individuals normalized around a mean of 0 on the abscissa (x) with ln(IgE) levels on the ordinate. Blue dots indicate subjects with eosinophil counts greater than the median for the MRCA panel.

provided further evidence that T-cell subsets do not have strong effects on these loci.

The variance (standard deviation) in our IgE-associated CpG loci was on average 4.4-fold larger in isolated eosinophils than in PBLs from the MRCA data set, indicating an attenuation of effect size in PBLs that would mask associations rather than magnify them. The power to detect cell-specific associations from PBLs depends on the proportion of each cell type, the effect size in specific cells, and the sample size. We estimated that we had 90% power to detect loci accounting for 10% of variance in IgE in the MRCA panel and >99% power in the combined panels (Extended Data Fig. 3). Although our ability to detect associations was enhanced by eosinophil counts that were above the normal range for many of our subjects (Table 1), the power estimates and the large observed effect size encourage epigenome-wide association studies of other diseases in which PBLs may be important.

We investigated the variance attributable to different loci in the MRCA panel through a stepwise regression that included all significant CGI associations together with differential white cell counts, age, sex and parental status. Using a cut-off of corrected $P < 0.1$ for inclusion, we found that *SLC25A33*, *LPCAT2* and *L2HGDH* predicted the serum IgE concentration independently of each other and of eosinophil counts.

In the MRCA panel, the top three CpGs independently explained 13.5% of IgE variation and the eosinophil counts explained a further 8.8%. In the SLSJ panel the top three CpGs explained 8.3% IgE variation and eosinophil counts a further 15.5%. (We were not able to estimate variances meaningfully in the PAPA data set, because the samples were selected by extreme IgE values). The regression models therefore matched the results from isolated eosinophils, leading us to the conclusion that the methylation status of eosinophils and their numbers were both related to IgE levels.

Methylation levels were highly correlated between loci, and similar estimations of variance were obtained with forced entry of other significantly associated markers, so the results do not imply that *SLC25A33*, *LPCAT2* and *L2HGDH* are the most important loci. DNA methylation is not meiotically heritable and the variance in IgE attributable to these loci does not impact on the problem of missing heritability.

Overall, the most significant association was to *cg01998785*, within a CGI adjacent to *LPCAT2*. *LPCAT2* encodes lyso-platelet-activating factor (PAF) acetyltransferase, which is essential to induced formation of PAF, a potent pro-inflammatory lipid mediator[24]. It is of interest that hypoactive variants of plasmatic PAF-acetylhydrolase are associated with atopy and asthma[25]. Other significant associations annotated to genes involved in phospholipid metabolism included lysoplasmalogenase (*TMEM86B*), *CEL* (carboxyl ester lipase) and *CLC*.

GATA1 is a known eosinophil transcription factor and subsequent investigations will determine if the other associated transcription factors *ZNF22*, *RB1* and *KLF1* regulate eosinophil activation. Other associations may encode proteins released from eosinophil granules, including

PRG2, *PRG3*, *SERPINC1* (antithrombin), *TFF1* (which may protect the mucosa), *CEL* and the polyvalent serine protease inhibitor *SPINK4*. Genes encoding mitochondrial proteins (*L2HGDH* and *SLC25A3*) are consistent with mitochondrial suppression of apoptosis in activated eosinophils[26].

Although the relationship between methylation and gene expression at these loci requires further investigation in isolated cells, our results support the recognition that eosinophils are an important source of cytokines and other pro-inflammatory molecules at the site of allergic inflammation[7]. Eosinophils are required locally for the maintenance of bone-marrow plasma cells[27], allowing direct regulation of IgE production in specialized environments. Clinically, the presence of eosinophilia in the peripheral blood or airways identifies a subgroup of refractory asthmatic individuals in whom therapies directed at eosinophils may be effective[28]. The measurement of methylation at these loci may further identify patients responsive to therapies directed at eosinophils or individual gene products.

Cigarette smoking may increase serum IgE, and we found anti-correlated associations to current cigarette smoking with *F2RL3* ($P = 8.6 \times 10^{-17}$) and *GPR15* ($P = 4.6 \times 10^{-9}$). The SLSJ data set confirmed these associations ($P = 2.5 \times 10^{-6}$ and $P = 6.6 \times 10^{-7}$), in keeping with previous studies[29,30]. Adjusting for smoking had minimal impact on the top hits for IgE and neither locus affected IgE in our subjects. *F2RL3* and *GPR15* may represent therapeutic targets to counter the effects of tobacco smoke and their methylation status may prospectively predict consequences of smoking.

Our epigenome-wide association study has discovered reproducible CGI associations accounting for a variation in the total serum IgE that is tenfold higher than that derived from large SNP genome-wide association studies[4]. In contrast to SNP studies, association to methylation levels captures responses to environmental factors and the loci should not be assumed to cause disease. Nevertheless, our findings suggest the presence of novel therapeutically tractable pathways underlying IgE production.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Holgate, S. T., Djukanovic, R., Casale, T. & Bousquet, J. Anti-immunoglobulin E treatment with omalizumab in allergic diseases: an update on anti-inflammatory activity and clinical efficacy. *Clin. Exp. Allergy* **35**, 408–416 (2005).
2. Busse, W. W. *et al.* Randomized trial of omalizumab (anti-IgE) for asthma in inner-city children. *N. Engl. J. Med.* **364**, 1005–1015 (2011).
3. Weidinger, S. *et al.* Genome-wide scan on total serum IgE levels identifies *FCER1A* as novel susceptibility locus. *PLoS Genet.* **4**, e1000166 (2008).
4. Moffatt, M. F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).

5. Levin, A. M. *et al.* A meta-analysis of genome-wide association studies for serum total IgE in diverse study populations. *J. Allergy Clin. Immunol.* **131,** 1176–1184 (2013).
6. Granada, M. *et al.* A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. *J. Allergy Clin. Immunol.* **129,** 840–845 (2012).
7. Nouri-Aria, K. T. *et al.* Cytokine expression during allergen-induced late nasal responses: IL-4 and IL-5 mRNA is expressed early (at 6 h) predominantly by eosinophils. *Clin. Exp. Allergy* **30,** 1709–1716 (2000).
8. Deaton, A. M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* **21,** 1074–1086 (2011).
9. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genet.* **38,** 1378–1385 (2006).
10. Löhning, M., Richter, A. & Radbruch, A. Cytokine memory of T helper lymphocytes. *Adv. Immunol.* **80,** 115–181 (2002).
11. Roessler, J. *et al.* Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. *BMC Res. Notes* **5,** 210 (2012).
12. Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448,** 470–473 (2007).
13. Ziyab, A. H. *et al.* Interactive effect of *STAT6* and *IL13* gene polymorphisms on eczema status: results from a longitudinal and a cross-sectional study. *BMC Med. Genet.* **14,** 67 (2013).
14. Laprise, C. The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population. *Genes Immun.* **15,** 247–255 (2014).
15. Palmer, L. J. *et al.* Independent inheritance of serum immunoglobulin E concentrations and airway responsiveness. *Am. J. Respir. Crit. Care Med.* **161,** 1836–1843 (2000).
16. Tykocinski, L. O. *et al.* A critical control element for interleukin-4 memory expression in T helper lymphocytes. *J. Biol. Chem.* **280,** 28177–28185 (2005).
17. Ansel, K. M., Djuretic, I., Tanasa, B. & Rao, A. Regulation of Th2 differentiation and *Il4* locus accessibility. *Annu. Rev. Immunol.* **24,** 607–656 (2006).
18. Relton, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* **41,** 161–176 (2012).
19. Lopez, A. F. *et al.* Recombinant human interleukin 5 is a selective activator of human eosinophil function. *J. Exp. Med.* **167,** 219–224 (1988).
20. Kita, H. Eosinophils: multifaceted biological properties and roles in health and disease. *Immunol. Rev.* **242,** 161–177 (2011).
21. Reinius, L. E. *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE* **7,** e41361 (2012).
22. Holt, P. G., Strickland, D. H., Wikstrom, M. E. & Jahnsen, F. L. Regulation of immunological homeostasis in the respiratory tract. *Nature Rev. Immunol.* **8,** 142–152 (2008).
23. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13,** 86 (2012).
24. Shindou, H. *et al.* A single enzyme catalyzes both platelet-activating factor production and membrane biogenesis of inflammatory cells. Cloning and characterization of acetyl-CoA:LYSO-PAF acetyltransferase. *J. Biol. Chem.* **282,** 6532–6539 (2007).
25. Kruse, S. *et al.* The Ile198Thr and Ala379Val variants of plasmatic PAF-acetylhydrolase impair catalytic activities and are associated with atopy and asthma. *Am. J. Hum. Genet.* **66,** 1522–1530 (2000).
26. Peachman, K. K., Lyles, D. S. & Bass, D. A. Mitochondria in eosinophils: functional role in apoptosis but not respiration. *Proc. Natl Acad. Sci. USA* **98,** 1717–1722 (2001).
27. Chu, V. T. *et al.* Eosinophils are required for the maintenance of plasma cells in the bone marrow. *Nature Immunol.* **12,** 151–159 (2011).
28. Pavord, I. D. *et al.* Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *Lancet* **380,** 651–659 (2012).
29. Breitling, L. P., Yang, R., Korn, B., Burwinkel, B. & Brenner, H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.* **88,** 450–457 (2011).
30. Wan, E. S. *et al.* Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum. Mol. Genet.* **21,** 3073–3082 (2012).

METHODS

Phenotyping. Ethical approval for the study was obtained from the NHS Multicentre Research Ethics Committee for the MRCA subjects; from the Swansea Joint Scientific Research Committee and Swansea Research Ethics Committee for the Swansea (PAPA) subjects; and from le Centre de Santé et des Services Sociaux de Chicoutimi for the SLSJ families. Written informed consent was obtained from all subjects or in the case of children, from their parents. Asthma was doctor defined. Following a standard respiratory questionnaire, all subjects submitted to venipuncture. Differential white cell counts were measured by automated counter. Total serum IgE and specific serum IgE to whole HDM (*Dermatophagoides pteronyssinus*) and Timothy grass pollen (*Phleum pratense)* were measured using the Immunocap FEIA (Pharmacia AB). The levels of specific IgE were converted to RAST units according to Pharmacia recommendations. A combined RAST index was calculated for each individual as the sum of the RAST scores to HDM and Timothy grass[15].

Detection of methylation status. DNA was extracted after red cell lysis and centrifugation to recover leukocyte nuclear pellets. DNA samples were bisulphite converted using the Zymo EZ DNA Methylation kit (Zymo Research) with an input of 1,000 ng. The assay was carried out as per the Illumina Infinium Methylation instructions, using the HumanMethylation27 BeadChips (Illumina). These interrogate 27,578 of CpG sites for the extent of DNA methylation. Data were visualized using the BeadStudio software, and samples that failed quality control were repeated. Raw methylation data were exported from the GenomeStudio software. For the Illumina HumanMethylation27 BeadChip data, quantile normalization of intensity was applied to all methylated and unmethylated probes for all samples together. The methylation β values were recalculated as the ratio of methylated probe signal/ (total signal + 100). The Touleimat and Tost[31] analysis pipeline was used for the HumanMethylation450 BeadChip. Individual data points with detection $P > 0.01$ or number of beads <3 were treated as missing data, as were samples with more than 20% missing probes. The lumi package[32] was used for background and colour bias correction. BeadChip ID and position on chip were included as categorical covariates to account for potential batch effects. Quantile normalization across samples was applied to probes within each functional category (CpG island, shelf, shore, and so on) separately to correct the shift of methylation β values between Infinium I and Infinium II probes on the HumanMethylation450 BeadChip. Probe overlaps with any frequent SNP (minimum allele frequency (MAF) $> 5\%$ in 1000 Genomes Project phase 1 EUR population) in the probe sequence or in position $+1$ or $+2$ of the query site (depending on Infinium I or Infinium II status) were removed. The use of meta-analysis to combine 27K and 450K data together with this implementation of the Tost pipeline ensured our analysis was not confounded by probe differences.

Isolation of human eosinophils. Isolation of human eosinophils was as described[33]. Briefly, platelet-rich plasma was removed from 200 ml using centrifugation, followed by Dextran-mediated sedimentation to remove erythrocytes and removal of mononuclear cells using a lymphocyte separation medium. Hypotonic lysis with sterile water removed remaining erythrocytes and other granulocytes were removed using negative selection with anti-CD16 MicroBeads. DNA was extracted using the QIAamp DNA Blood Mini Kit. Methylation was assessed using Illumina 450K arrays, with analysis restricted to significantly associated probes from the meta-analysis.

Statistical analyses. To investigate the association with the total serum IgE concentration we tested for association with log-normalized IgE (ln(IgE)) as response with methylation (β) at each locus as predictor while including sex, age, parent indicator, age × sex and age × parent interactions in the model, together with batch indicators captured by Illumina chip ID and position of chip (such as operators, sample wells, plates, runs, and reagents). We applied inverse normal transformation to methylation measures to remove the effect of outliers. We used the R function lme() in the nlme package to implement a linear mixed model, assuming a compound symmetry variance–covariance structure to account for correlation of phenotypes among family members. The R code for the discovery stage of association in the MRCA panel was:

index = !is.na(methylation)
fam = familyID[index]
par = parent[index]
methylation = methylation [index]
methylation = qnorm(rank(methylation)/(length(methylation)+1),mean = 0,sd = 1)
lnige = LNIGE[index]
age = AGE[index]
sex = SEX[index]
lm2 = lme(lnige~sex+age+methylation+par+sex*age+age*par,random = ~1 |fam)

The residual methylation values after removal of the effects of chip ID and position for the genome-wide significant loci in the MRCA, PAPA and SLSJ panels are provided in Supplementary Tables 3–5, together with phenotypic and covariate parameters. We calculated FDRs and applied Bonferroni corrections to adjust for multiple comparisons to 27,578 probes. The same analyses were carried out in the PAPA and SLSJ subjects before meta-analysis of the three data sets. We use a weighted *z*-score method for meta-analysis based on *P* value and effect direction from individual studies with weights proportion to the square root of the sample size of the individual study[34]. SNPs and indels from the 1000 Genomes Project phase 1 release (14 March 2012 haplotypes) were imputed using MINIMAC[35]. SNPs or indels with imputation quality score $R^2 < 0.3$ were removed from downstream analysis. We carried out Mendelian randomization to assess the causal effect of IL-4 methylation on IgE level through a two-stage least squares instrumental variable regression[36] implemented in the ivreg2.r program (http://diffuseprior. wordpress.com/2012/05/03/an-ivreg2-function-for-r/). We tested association trends in isolated eosinophils by exact regression (Cytel Studio 9) with asthma/ high IgE coded as 2, asthma/low IgE coded as 1, and controls as 0. Covariates for age, sex and batch were included in the model and to test the hypothesis that low levels of methylation were associated with high IgE, *P* values were one-sided. Differences in methylation between peripheral blood leukocyte subsets were assessed with Kruskal–Wallis tests, using two-sided *P* values.

31. Touleimat, N. & Tost, J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* **4,** 325–341 (2012).
32. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24,** 1547–1548 (2008).
33. Ferland, C. *et al.* Eotaxin promotes eosinophil transmigration via the activation of the plasminogen-plasmin system. *J. Leukoc. Biol.* **69,** 772–778 (2001).
34. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
35. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through prephasing. *Nature Genet.* **44,** 955–959 (2012).
36. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27,** 1133–1163 (2008).
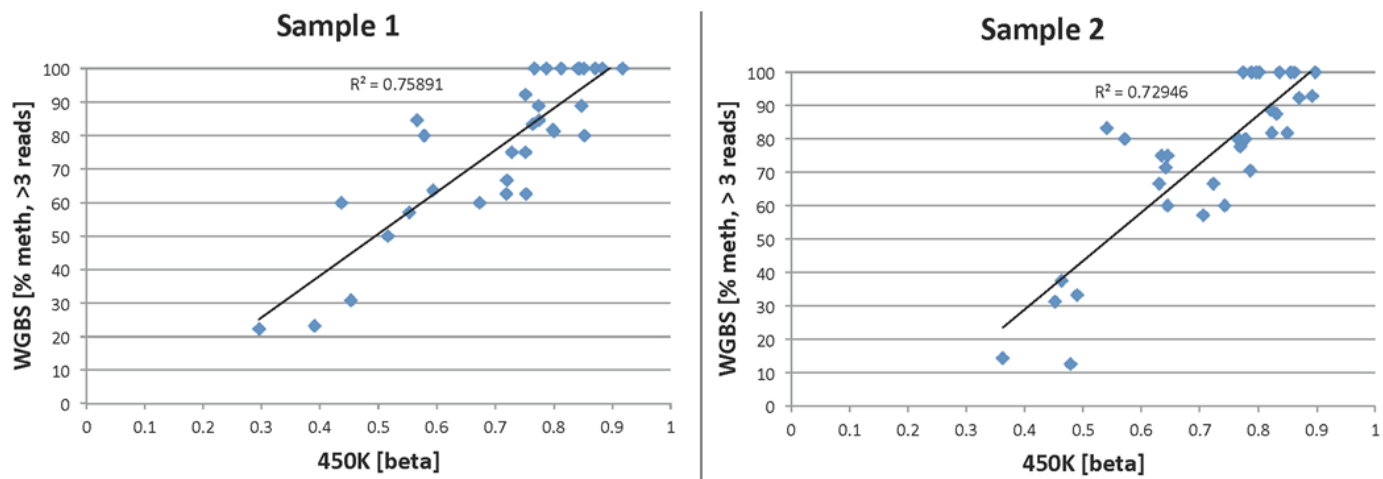
Extended Data Figure 1 | Concordance in methylation status at IgE-associated loci when comparing whole-genome bisulphite sequencing with the Illumina platform. These results were produced by us (E.G. and T.M.P.) at the Genome Québec Innovation Centre. The figures show a comparison between IgE-associated CpG probes using Illumina 450K (x-axis) and whole-genome bisulphite sequencing (WGBS) (y-axis) platforms for two samples (left and right panels) with 20-fold sequence coverage. The results show a high R^2 between platforms (0.76 and 0.73). The median of the correlation coefficients for our IgE-associated loci across 30 different samples (using WGBS at various depths) was $R^2 = 0.76$. This can be compared with the global assessment of all overlapping 450K sites, which is $R^2 = 0.81$.
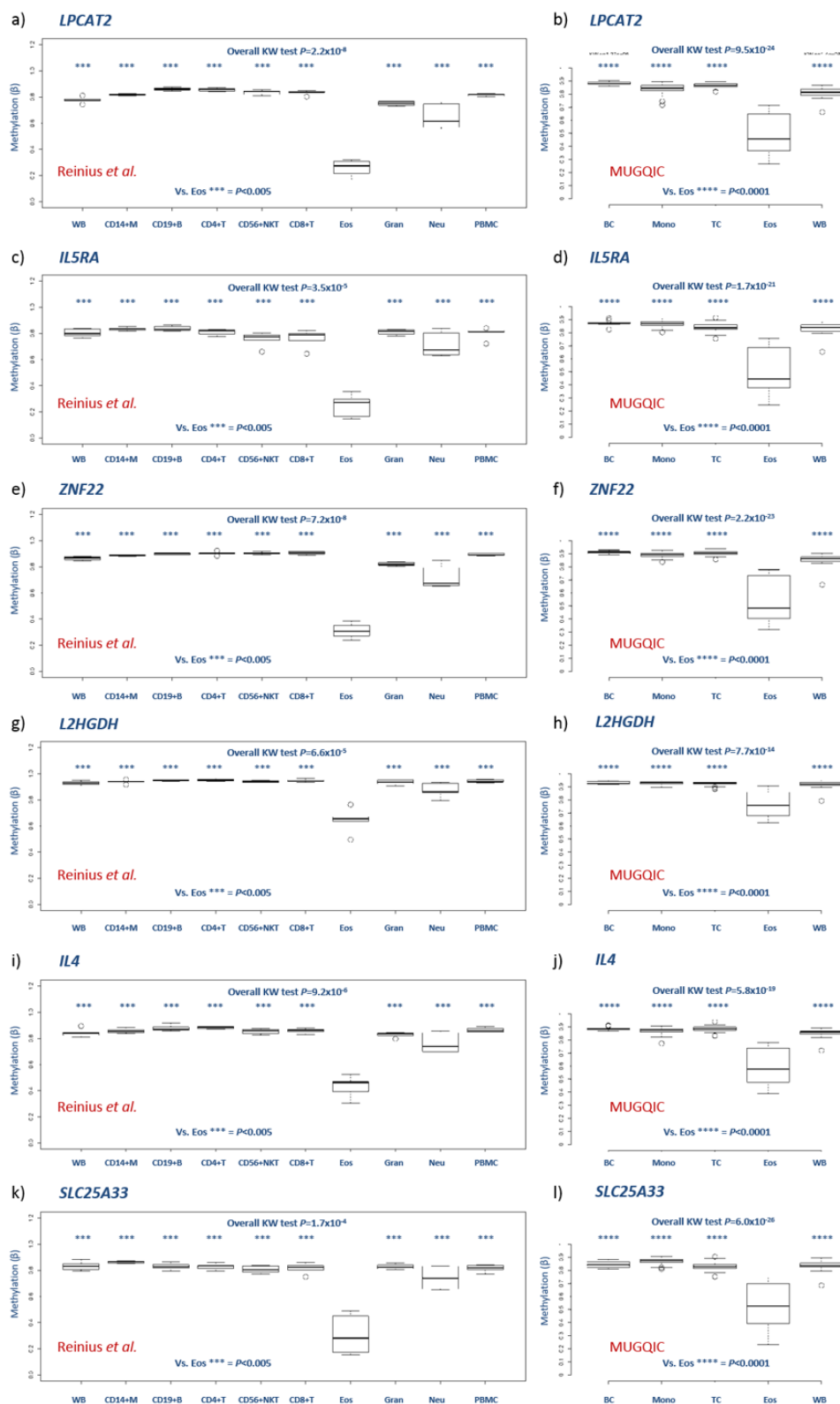
Extended Data Figure 2 | Distribution of methylation status at IgE-associated loci in isolated leukocyte subsets. a–l, The figure shows the distribution of methylation in PBL subsets at the most strongly IgE-associated loci. CpG methylation was measured by the Illumina Infinium 450K platform. Boxplots show means and interquartile ranges. **a, c, e, g, i, k**, Results from publically available data derived from six healthy controls[21]. Lower levels of methylation with wider variation are observed in eosinophils when compared to whole blood (WB) and subsets comprising CD14+ monocytes (CD14+M); CD19+ B cells (CD19+B); CD4+ T cells (CD4+T); CD56+ natural killer cells (CD56+NK); CD8+ T cells (CD8+T); granulocytes (Gran); neutrophils (Neu) and peripheral blood mononuclear cells (PBMCs). **b, d, f, h, j, l**, Results from cells isolated and analysed by us at the McGill University Genome Québec Innovation Centre (MUGQIC). Eosinophils (Eos) (from 24 subjects in the SLSJ panel) also show lower levels of methylation with wider variation compared to whole blood (22 SLSJ subjects), and to subsets including B cells (BC; 9 control subjects), monocytes (Mono; 76 control subjects) and T cells (TC; 74 control subjects). KW Test, Kruskal–Wallis one-way analysis of variance.
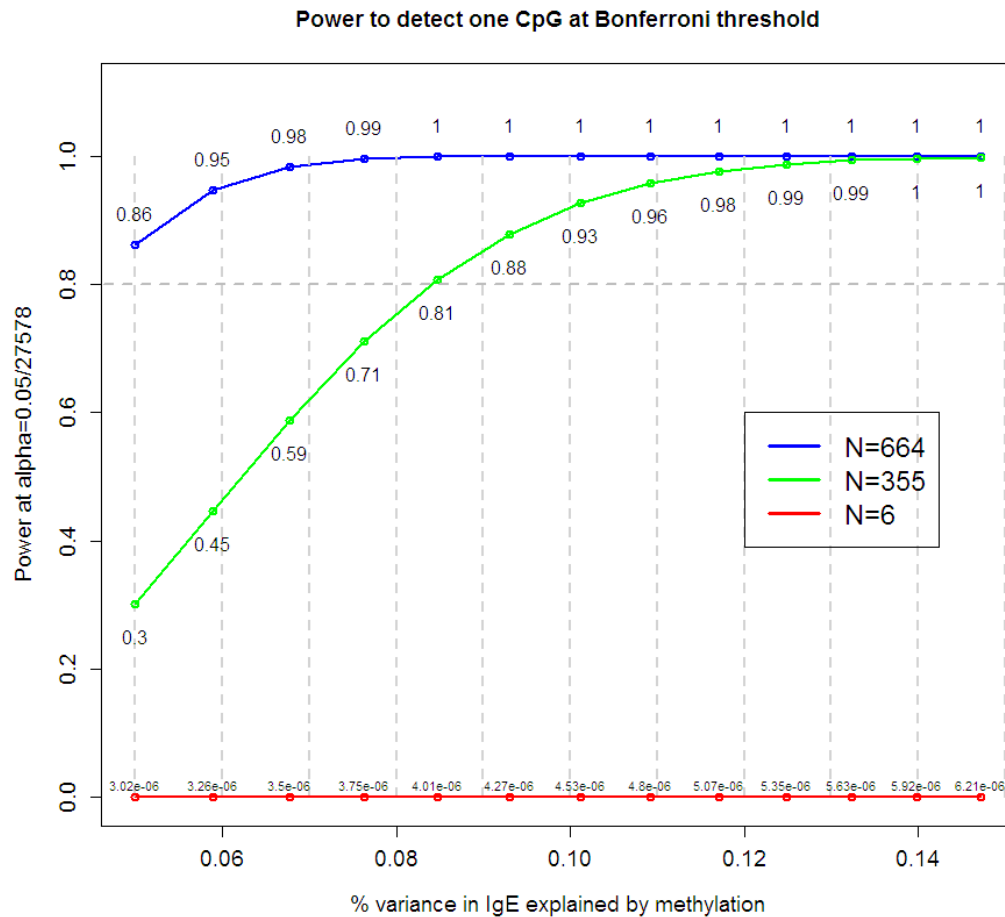
Extended Data Figure 3 | Power estimations to detect eosinophil-specific effects in DNA from PBL. The figure shows that our original MRCA data set (green line) and our combined data set (blue line) are well powered to detect signals of the magnitude observed in our three groups of subjects. The red line shows the power of a sample size of six, as described previously[21], to detect differences in CpG methylation in unfractionated PBLs. The mean variance (as standard deviation (s.d.)) for the IgE-associated loci was 0.036 in PBLs from our primary MRCA panel and 0.023 in the whole blood normal samples from ref. 21, demonstrating that our results were consistent with the previous experiment.

Extended Data Table 1 | Comparison of surrogate variable analyses with direct white cell counts in association models

					a) Before	b) Adjusting for Houseman cell proportions							c) Adjusting for white cell counts					
Chr.	Position	Symbol	probe	pMethy	pMethy	pCD8T	pCD4T	pNK	pBCELL	pMON	pGRAN	pMethy	pEOS	pNEU	pLYM	pMON	pBAS
1	9521654	SLC25A33	cg18783781	4.96E-14	5.58E-13	0.3625663	0.6489145	0.0830001	0.0735391	0.7715233	0.2564499	6.14E-06	0.00065	0.866661	0.678896	0.561016	0.551752
16	54100210	LPCAT2	cg01998785	1.24E-13	3.81E-13	0.1560292	0.3633042	0.034881	0.0300831	0.7239839	0.1881862	2.36E-06	0.0003	0.294806	0.300248	0.445243	0.785135
14	49849865	L2HGDH	cg15996947	7.39E-13	3.81E-11	0.3709841	0.6823007	0.125117	0.1583239	0.9387673	0.3768824	3.21E-05	0.000295	0.38062	0.519495	0.662394	0.890493
10	44815441	ZNF22	cg01614759	4.42E-12	9.09E-11	0.3526504	0.6420391	0.1382946	0.0810812	0.654088	0.3845248	7.61E-05	0.000216	0.305013	0.42021	0.601202	0.937252
3	3127530	IL5RA	cg10159529	5.06E-12	1.70E-10	0.8786628	0.9388517	0.2941882	0.2701973	0.9752202	0.5129874	0.000104	6.81E-05	0.877227	0.802458	0.551635	0.756105
1	172153108	SERPINC1	cg01770400	6.59E-12	1.54E-11	0.1182883	0.3063038	0.0323251	0.0368702	0.574556	0.1896	6.36E-06	0.000109	0.236406	0.186757	0.667289	0.90807
21	42659768	TFF1	cg02643667	7.87E-12	6.81E-11	0.234861	0.4646176	0.0330799	0.0706689	0.797703	0.1753888	9.07E-06	7.18E-06	0.384611	0.725574	0.672431	0.976409
5	132036424	IL4	cg26787239	1.59E-11	2.13E-10	0.5058003	0.9922092	0.1723318	0.1996409	0.7364825	0.6544331	1.25E-05	1.66E-05	0.269226	0.259295	0.614387	0.77939
2	110014086	LIMS3	cg18879041	4.06E-11	2.57E-09	0.3708111	0.7030526	0.1613161	0.2166898	0.9745824	0.474026	0.000865	0.000256	0.280498	0.382954	0.500203	0.963467
13	47774920	RB1	cg13221796	5.66E-10	9.26E-11	0.1278218	0.3641314	0.0353766	0.0458363	0.7786051	0.2987393	2.84E-06	1.83E-05	0.029037	0.081617	0.516731	0.982861
19	60432000	TMEM86B	cg26457013	1.02E-09	3.14E-08	0.3305229	0.5920938	0.036091	0.0792481	0.9078355	0.2711535	0.000691	1.35E-05	0.363948	0.317448	0.558421	0.740734
9	100745613	COL15A1	cg20503329	1.19E-09	3.14E-08	0.4772819	0.7388277	0.1359031	0.1343368	0.9433195	0.4284566	0.001341	1.17E-05	0.443838	0.49736	0.401174	0.959364
1	31166502	SDC3	cg07689731	2.33E-09	4.79E-09	0.5074834	0.7985815	0.0774502	0.0726484	0.7924958	0.4108301	0.002906	1.48E-05	0.415182	0.483806	0.437372	0.932638
21	45317773	ADARB1	cg09676390	1.21E-08	1.32E-07	0.43369	0.6702876	0.0791369	0.1103603	0.9234331	0.3292778	0.000472	3.48E-07	0.432091	0.530921	0.462608	0.820656
9	134926722	CEL	cg03693099	1.84E-08	1.01E-07	0.2926779	0.512945	0.0818301	0.1071869	0.9754856	0.3848947	0.000383	9.27E-06	0.099207	0.18224	0.575481	0.93957
19	12859426	KLF1	cg26136776	3.34E-08	5.42E-07	0.2328607	0.4764967	0.0498737	0.087521	0.7661611	0.2312155	0.008494	4.25E-06	0.308908	0.378613	0.526765	0.764392
X	48529968	GATA1	cg00536175	7.85E-08	5.70E-07	0.6056632	0.9159691	0.163687	0.1102612	0.8082841	0.5675393	0.002444	1.31E-06	0.230592	0.333849	0.423511	0.745501
12	10222881	TMEM52B	cg25494227	1.29E-07	2.79E-07	0.4810187	0.6746518	0.0457622	0.0735827	0.9017706	0.3386288	8.31E-05	2.05E-08	0.349366	0.362883	0.456694	0.75634
9	33229641	SPINK4	cg00079056	1.54E-07	1.23E-06	0.3893041	0.6595419	0.067607	0.090288	0.950103	0.3262354	0.015058	9.70E-07	0.327361	0.481372	0.49343	0.779557
12	15017287	PDE6H	cg09447105	2.16E-07	1.80E-06	0.6038815	0.9752817	0.1325361	0.2619219	0.6725442	0.6818338	0.002327	6.73E-07	0.185319	0.308805	0.650588	0.72691
16	4778723	FLJ25410	cg05215575	3.11E-07	7.40E-08	0.6354446	0.6693211	0.0382164	0.0635217	0.8705757	0.2084314	0.011151	9.08E-08	0.699683	0.891282	0.341595	0.856081
7	142369547	KEL	cg17784922	4.24E-07	9.07E-07	0.6926023	0.9854571	0.0699431	0.1577229	0.8028496	0.5052827	0.001391	3.21E-08	0.505128	0.698535	0.367309	0.99912
7	29199993	CHN2	cg21472642	6.51E-07	7.39E-07	0.5225896	0.8312874	0.0708089	0.0764603	0.9322296	0.4725488	8.45E-06	3.38E-09	0.104506	0.231601	0.595919	0.815002
2	191009029	FLJ20160	cg15998761	9.33E-07	9.09E-06	0.4236314	0.824669	0.1393183	0.1145142	0.8527446	0.4643437	0.043732	5.71E-07	0.331558	0.449501	0.418024	0.785659
6	25862169	SLC17A4	cg21627181	1.14E-06	1.42E-06	0.7663137	0.9149808	0.0591244	0.1414494	0.9511792	0.3972344	0.025202	5.30E-08	0.609342	0.865393	0.467443	0.746032
14	49849874	L2HGDH	cg20189937	1.32E-06	1.16E-05	0.503935	0.9204867	0.0921306	0.1950088	0.8817908	0.5174749	0.007269	1.26E-07	0.297519	0.525586	0.480584	0.958405
17	39822093	ITGA2B	cg17749520	1.37E-06	9.24E-06	0.3677214	0.6752543	0.046072	0.1157269	0.8717913	0.3274826	0.009625	3.11E-07	0.286277	0.406727	0.666483	0.903995
17	39701014	SLC4A1	cg03580247	1.51E-06	1.76E-05	0.4966594	0.7030166	0.1566437	0.1201467	0.7706213	0.4809553	9.19E-05	7.97E-09	0.262889	0.282218	0.522557	0.944193
20	41789039	FAM112A	cg11398517	2.42E-06	2.11E-05	0.431888	0.7316467	0.0506171	0.209415	0.8337169	0.4256838	0.006763	1.86E-07	0.220078	0.444385	0.592602	0.918166
11	56904791	PRG3	cg24459209	3.33E-06	1.29E-05	0.5460024	0.9366639	0.0692143	0.1545826	0.8947187	0.4338765	0.094895	2.69E-07	0.444762	0.62114	0.500854	0.993962
12	7792928	CLEC4C	cg22194129	5.46E-06	0.0001208	0.4977092	0.8378959	0.0710935	0.6019842	0.8886589	0.4447889	0.023532	2.38E-08	0.408481	0.759002	0.521453	0.818552
14	95011802	C14orf49	cg16522484	8.04E-06	0.0001668	0.4308412	0.7282314	0.1486239	0.1112817	0.9482318	0.4206034	0.034072	1.01E-07	0.285777	0.449644	0.49125	0.837548
X	48529554	GATA1	cg22543648	9.51E-06	7.12E-05	0.682124	0.9511741	0.1312401	0.1917101	0.6997158	0.4748147	0.02664	2.00E-08	0.585973	0.689124	0.425854	0.888172
2	11727816	NTSR2	cg25657834	1.02E-05	2.04E-05	0.9354409	0.6331433	0.1373482	0.3261244	0.4920064	0.8015518	5.67E-05	4.84E-10	0.191813	0.526463	0.689339	0.954103
19	44912054	IRGC	cg26251865	1.76E-05	9.19E-06	0.714389	0.9562959	0.0555475	0.1095196	0.6922884	0.4367507	0.011533	1.35E-06	0.495745	0.567965	0.468957	0.977824
19	44919789	CLC	cg18254848	1.77E-05	2.33E-05	0.7122613	0.874791	0.0834931	0.1954486	0.6341788	0.5277066	0.161753	4.77E-08	0.462599	0.67872	0.424877	0.937408
6	10635801	GCNT2	cg26385286	1.78E-05	3.88E-05	0.5747779	0.9758809	0.0654066	0.2046998	0.6574459	0.5308099	0.070861	4.31E-08	0.339931	0.593173	0.451377	0.902264
3	150064527	CPA3	cg13424229	1.89E-05	2.29E-06	0.4329161	0.6962458	0.1048729	0.0674688	0.7084296	0.6214492	0.001236	4.96E-07	0.054043	0.094343	0.451245	0.955257

a–c, The table shows the P values from regression models in the MRCA panel that predict ln(IgE) at each locus before adjusting cell counts (**a**); after adjusting cell counts using Houseman surrogate variables (**b**); and after adjusting for cell subsets counted in our data (**c**). pMethyl measures the strength of association to IgE. CD8, CD4 and NK are T-cell subsets. All models are adjusted for sex, age, methylation, parent/child status, sex × age interaction, and age × parent interaction. BAS, basophils; EOS, eosinophils; GRAN, granulocytes; LYM, lymphocytes; MON, monocytes; NEU, neutrophil.

LETTER

A circuit mechanism for differentiating positive and negative associations

Praneeth Namburi[1,2]*, Anna Beyeler[1]*, Suzuko Yorozu[3], Gwendolyn G. Calhoon[1], Sarah A. Halbert[1,4], Romy Wichmann[1], Stephanie S. Holden[1,5], Kim L. Mertens[1,6], Melodi Anahtar[1,5], Ada C. Felix-Ortiz[1,2], Ian R. Wickersham[7], Jesse M. Gray[3] & Kay M. Tye[1]

The ability to differentiate stimuli predicting positive or negative outcomes is critical for survival, and perturbations of emotional processing underlie many psychiatric disease states. Synaptic plasticity in the basolateral amygdala complex (BLA) mediates the acquisition of associative memories, both positive[1,2] and negative[3–7]. Different populations of BLA neurons may encode fearful or rewarding associations[8–10], but the identifying features of these populations and the synaptic mechanisms of differentiating positive and negative emotional valence have remained unknown. Here we show that BLA neurons projecting to the nucleus accumbens (NAc projectors) or the centromedial amygdala (CeM projectors) undergo opposing synaptic changes following fear or reward conditioning. We find that photostimulation of NAc projectors supports positive reinforcement while photostimulation of CeM projectors mediates negative reinforcement. Photoinhibition of CeM projectors impairs fear conditioning and enhances reward conditioning. We characterize these functionally distinct neuronal populations by comparing their electrophysiological, morphological and genetic features. Overall, we provide a mechanistic explanation for the representation of positive and negative associations within the amygdala.

The BLA, including lateral and basal nuclei of the amygdala[11], receives sensory information from multiple modalities[12–14], and encodes motivationally relevant stimuli[15–17]. Partially non-overlapping populations of BLA neurons encode cues associated with appetitive or aversive outcomes[8,9]. The acquisition of the association between a neutral stimulus and an aversive outcome such as a foot shock has been shown to induce long term potentiation (LTP) of synapses onto lateral amygdala neurons[3,4], mediated by postsynaptic increases in α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPAR)-mediated currents[5,18] in a N-methyl-D-aspartate receptor (NMDAR)-dependent manner[19,20]. Similarly, increases in glutamatergic synaptic strength of inputs to BLA neurons are necessary for the formation of a stimulus–reward association[1]. Yet the similarity in neural encoding and synaptic changes induced by learning a positive or negative association and the contrasting nature of the ensuing outputs (reward-seeking or fear-related behaviours) presents an ostensible paradox: how is it possible that potentiation of synapses onto neurons in the BLA can underlie learned associations that lead to such different behavioural responses?

One hypothesis is that BLA neurons project to many downstream regions, including the canonical circuits for reward and fear[14,30], and the neurons that project to different targets undergo distinct synaptic changes with positive or negative associative learning. For example, BLA projections to the NAc have been implicated in reward-related behaviours[16,21,22], while BLA projections to the CeM have been linked to the expression of conditioned fear[23–25]. However, the unique syn-

aptic changes onto projection-identified BLA neurons have never been explored.

To investigate this, we selected the NAc and CeM as candidate target regions and examined the synaptic changes onto either NAc-projecting BLA neurons (NAc projectors) or CeM-projecting BLA neurons (CeM projectors) following fear conditioning or reward conditioning (Fig. 1). To identify the projection target of BLA neurons, we injected retrogradely travelling fluorescent beads (retrobeads) into either the NAc or CeM to label BLA neurons projecting axon terminals to these regions (Fig. 1a and Extended Data Fig. 1). After retrobead migration upstream to BLA cell bodies, we trained mice in fear or reward conditioning paradigms wherein a tone was paired with either a foot shock or sucrose delivery. Mice in reward conditioning groups were food restricted 1 day before the conditioning session to increase motivation to seek sucrose (Extended Data Fig. 1). AMPAR/NMDAR ratio, a proxy for glutamatergic synaptic strength, increases after either fear or reward conditioning in the BLA[1,2,5,18]. We used matched experimental parameters across groups in an acute slice preparation stimulating axons arriving via the internal capsule and performing whole-cell patch-clamp recordings in retrobead-identified NAc projectors and CeM projectors, which we observed to be topographically intermingled (Fig. 1b and Extended Data Fig. 2).

We found that in NAc projectors, fear conditioning decreased the AMPAR/NMDAR ratio relative to controls exposed to the same number of tones and shocks, but where the tones and shocks were unpaired (Fig. 1c). Conversely, following the acquisition of the association between a tone and sucrose delivery, synapses on NAc projectors showed an increase in AMPAR/NMDAR ratio relative to unpaired controls that were also food restricted and received the same number of tones and volume of sucrose (Fig. 1d). Importantly, we also included naive and food-restricted naive groups (Fig. 1c–f), as food restriction itself could alter AMPAR/NMDAR ratio (Extended Data Fig. 3).

In contrast, synapses on CeM projectors from the paired group showed an increase in AMPAR/NMDAR ratio following fear conditioning, relative to unpaired controls (Fig. 1e). Following reward conditioning, CeM projectors from mice that learned the tone–sucrose association showed a decrease in AMPAR/NMDAR ratio relative to unpaired controls (Fig. 1f). In addition to AMPAR/NMDAR ratios, we also examined paired-pulse ratios, and did not detect any differences between groups (Extended Data Fig. 3), suggesting a postsynaptic mechanism of plasticity.

These results support a model wherein NAc and CeM projectors undergo opposing changes in synaptic strength following fear and reward learning, such that relative synaptic strengths onto CeM projectors increase following fear conditioning and decrease following reward learning. Conversely, relative synaptic strengths onto NAc

[1]The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Neuroscience Graduate Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, NRB 356, Boston, Massachusetts 02115, USA. [4]Undergraduate Program in Neuroscience, Wellesley College, Wellesley, Massachusetts 02481, USA. [5]Undergraduate Program in Neuroscience, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [6]Master's Program in Biomedical Sciences, University of Amsterdam, Amsterdam 1098 XH, The Netherlands. [7]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.
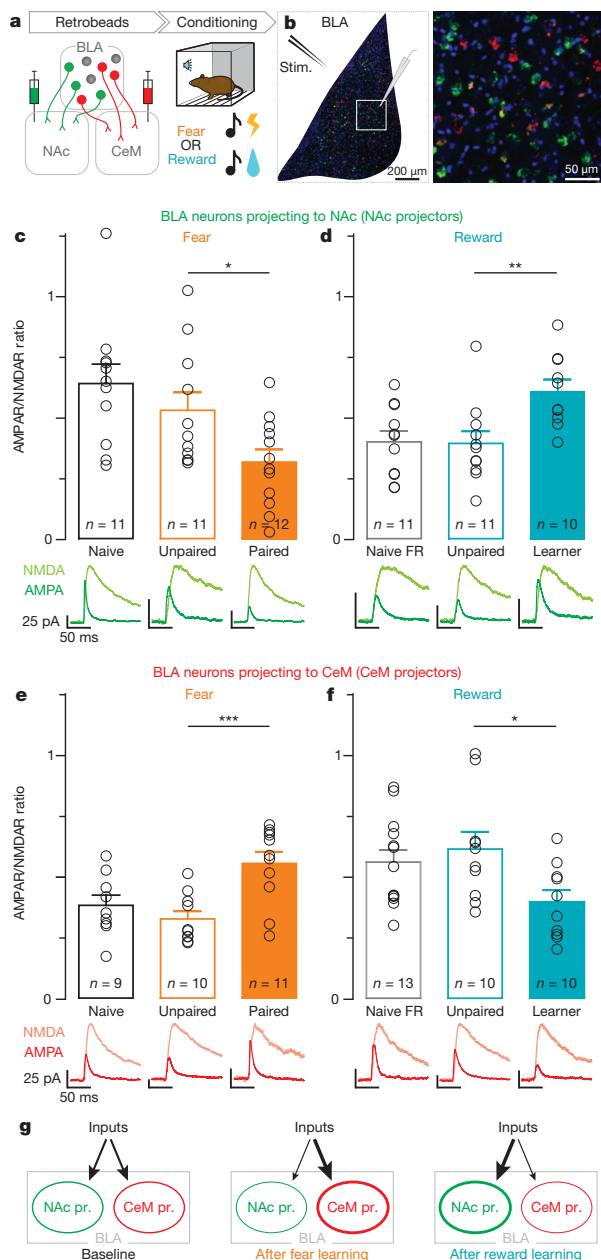
Figure 1 | Opposite changes in AMPAR/NMDAR following fear or reward conditioning in BLA neurons projecting to NAc or CeM. **a**, After injecting retrobeads into NAc or CeM, animals underwent either fear or reward conditioning. **b**, Confocal image of retrobead-labelled BLA neurons, with schematic of stimulation and recording sites (left); region in white square is enlarged (right). DAPI is shown in blue. **c–f**, One-way ANOVAs were performed on AMPAR/NMDAR ratios collected after conditioning. Open circles reflect individual data points, number of neurons are shown in each bar and representative traces for each group are below the bar. Results show mean and s.e.m. FR, food restricted. **c**, AMPAR/NMDAR ratio was related to training condition ($F_{2,33} = 5.844$, $P = 0.0070$) and significantly lower in the paired fear group relative to the unpaired fear group ($t_{31} = 2.21$, $*P < 0.05$). **d**, AMPAR/NMDAR ratio was related to training condition ($F_{2,31} = 6.53$, $P = 0.0046$) and reward learners showed a greater AMPAR/NMDAR ratio than mice in the unpaired reward group ($t_{29} = 3.20$, $**P < 0.01$). **e**, In CeM projectors, AMPAR/NMDAR ratio was related to fear conditioning ($F_{2,29} = 8.72$, $P = 0.0012$) and was greater in the paired group relative to the unpaired group ($t_{27} = 3.99$, $***P < 0.001$). **f**, In CeM projectors, AMPAR/NMDAR ratio was altered by reward learning ($F_{2,32} = 3.63$, $P = 0.039$), and was lower in learners relative to the unpaired group ($t_{30} = 2.57$, $*P < 0.05$). **g**, Proposed model, arrow thickness represents relative synaptic strength. CeM pr., CeM projectors; NAc pr., NAc projectors.
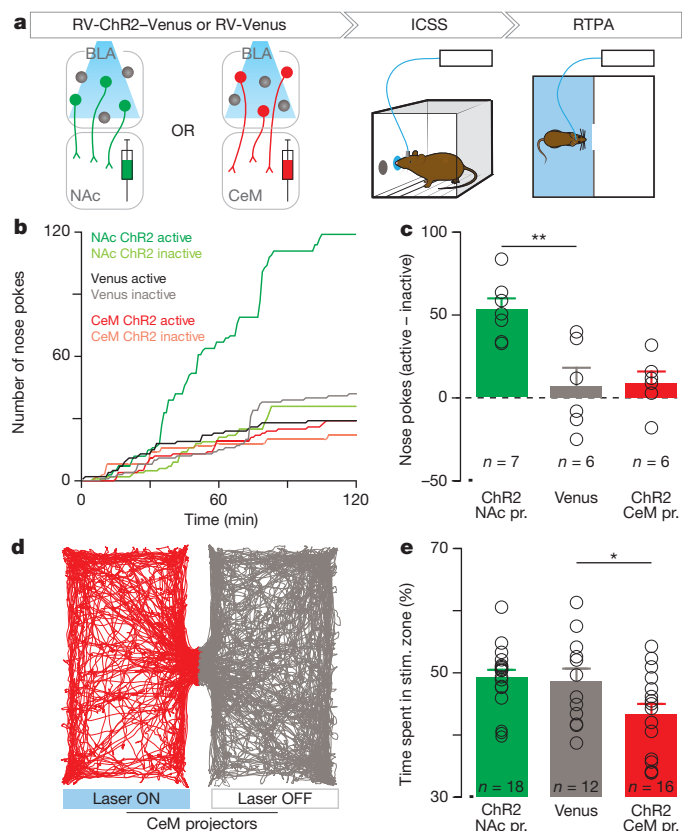
Figure 2 | Within the BLA, photostimulation of NAc or CeM projectors causes positive or negative reinforcement, respectively. **a**, After rabies virus injection into NAc or CeM, animals were tested using intracranial self-stimulation (ICSS) and real-time place avoidance (RTPA) assays. **b**, Representative traces of nose poke responses during ICSS. **c**, The relative number of active nose pokes was related to the experimental group (one-way ANOVA, $F_{2,18} = 10.50$, $P = 0.0012$), and was significantly increased by photostimulation of NAc projectors (NAc pr.) in comparison to controls ($t_{16} = 4.00$, $**P < 0.01$). CeM pr., CeM projectors. **d**, Representative locomotor trace from an animal receiving CeM projector photostimulation during RTPA. **e**, The percentage of time spent in the photostimulation-paired zone was related to the experimental group (one-way ANOVA, $F_{2,45} = 4.38$, $P = 0.019$) and was significantly decreased by photostimulation of CeM projectors in comparison to controls ($t_{43} = 2.25$, $*P < 0.05$). Results show mean and s.e.m.

projectors decrease following fear conditioning and increase following reward learning (Fig. 1g). However, these findings raise new questions. First, is there a causal relationship between the activity of BLA neurons projecting to the NAc and reward-related behaviours, and between the activity of CeM projectors and fearful or aversive behaviours? Second, what defining features of NAc and CeM projectors might endow them with their opposing functions?

To test whether there is a causal relationship between populations of projection-identified BLA neurons and behaviour, we first used a retrogradely infectious rabies viral (RV) vector[26] to express channelrhodopsin-2 (ChR2) fused to a fluorescent reporter (Venus), or a control virus carrying Venus alone (RV-Venus) in BLA neurons projecting to either the NAc or the CeM (Fig. 2a).

Following verification of functional ChR2 expression and retrograde transport from either the NAc or CeM back to the BLA (Extended Data Fig. 7), we tested animals receiving injections of RV-ChR2–Venus or RV-Venus into either the NAc or CeM and implantation of an optical fibre over the BLA on an intracranial self-stimulation (ICSS) task (Extended Data Fig. 4). Consistent with previous reports that photostimulation of BLA axons in the NAc produced ICSS[21], we observed ICSS upon photostimulation of BLA cell bodies projecting to NAc (Fig. 2b, c). Given that we could not elicit robust nose-poke responses

for CeM projector photostimulation, we next tested CeM projectors in a closed-loop real-time place avoidance assay (RTPA), where an animal freely explored two chambers, one in which the mouse received photostimulation of CeM projectors. Photostimulation of CeM projectors caused robust avoidance of the light-paired side (Fig. 2d, e). Consistent with our data and previous studies, we demonstrated a causal relationship between NAc projectors and positive reinforcement and CeM projectors and negative reinforcement (avoidance).

We went on to probe the necessity of NAc or CeM projectors in mediating reward or fear conditioning. The acquisition of fear[19] and reward[1] associations is mediated by an NMDAR-dependent LTP mechanism thought to require simultaneous glutamate release and postsynaptic depolarization. Thus, we tested whether projection-specific hyperpolarization during the presentation of the unconditioned stimulus could impair learning in a valence-specific manner.

To this end, we bilaterally infused an adeno-associated viral vector carrying halorhodopsin fused to an enhanced yellow fluorescent protein, or a no-opsin control (eYFP only), expressed in a Cre-dependent manner (double-floxed inverted open reading frame) into the BLA (Extended Data Fig. 5). We then bilaterally infused a retrogradely travelling canine adenovirus carrying Cre recombinase (CAV-Cre) into either the NAc or CeM (Fig. 3a). We illuminated the BLA with yellow light only during conditioned–unconditioned stimulus pairing, that is, during shock or sucrose consumption. Photoinhibition of CeM projectors during the conditioned–unconditioned stimulus pairing impaired conditioned freezing (Fig. 3b and Extended Data Fig. 6) and enhanced conditioned reward seeking (Fig. 3c).

Next, because there was no topographical separation between NAc and CeM projectors (Fig. 1b), which are both glutamatergic[11,21,27,28], we searched for distinguishing characteristics of these functionally distinct neuronal populations. As the BLA is known to have some heterogeneity

in electrophysiological and morphological characteristics[11,29], we compared these features between NAc and CeM projectors. While we did not observe differences in action potential half-width (Fig. 4a, b), threshold to spike (Fig. 4c, d), or intrinsic excitability (Fig. 4e, f), we did observe a significant difference in action potential accommodation (Fig. 4g and Extended Data Fig. 7). To investigate the morphological

Figure 4 | Electrophysiological, morphological and transcriptional profiles of NAc and CeM projectors. a, Population average of action potential traces. CeM pr., CeM projectors; NAc pr., NAc projectors. b, No detectable difference in action potential half-width (unpaired t-test, $t_{20} = 1.82$, $P = 0.085$). Open circles represent individual data points. c, Representative trace from action potential threshold detection protocol. d, No detectable difference in action potential threshold between NAc and CeM projectors (unpaired t-test, $t_{20} = 1.05$, $P = 0.31$). e–g, Representative trace (e) from current injection protocol to determine firing rate responses (f) and action potential probability (g) over time, which was different between NAc and CeM projectors (interaction, two-way ANOVA $F_{9,180} = 2.32$, $P = 0.017$) in the first 100 ms of current injection ($t_{200} = 4.55$, ***$P < 0.001$). h, Representative reconstructions showing dendritic branching pattern in the coronal plane. Cells were classified into pyramidal or stellate based on the presence of an apical tuft (right). i, Sholl analysis of neuron reconstructions. j, Schematic of transcriptome profiling. k, Candidate genes identified as differentially expressed between NAc and CeM projectors at a 0.01 quantile fold-change threshold, corresponding to a false discovery rate (FDR) of 26.2% (see also Extended Data Fig. 9) across two independent repetitions of RNA-seq ($n=8$ mice for NAc pr., $n=9$ mice for CeM pr. total). Results show mean and s.e.m.

Figure 3 | Photoinhibition of CeM projectors impairs fear learning and enhances reward learning. a, Halorhodopsin (NpHR) was expressed bilaterally either in NAc- or CeM-projecting BLA neurons using a dual-virus recombination strategy. Mice underwent fear or reward conditioning and yellow light was delivered to the BLA during each unconditioned stimulus. b, Time course of percentage freezing and average freezing in trials 6–8 (inset). Average freezing was related to experimental condition (one-way ANOVA, $F_{2,40} = 6.68$, $P = 0.0033$) and was significantly reduced by photoinhibition of CeM projectors, relative to controls ($t_{38} = 3.46$, **$P < 0.01$; see inset). c, Time course of normalized number of port entries relative to cue presentation during reward conditioning and average number of normalized port entries (<8 s latency, inset). z-score of port entries was related to the experimental condition (one-way ANOVA, $F_{2,31} = 9.23$, $P = 0.0008$) and was significantly increased by photoinhibition of CeM projectors, relative to controls ($t_{29} = 4.11$, ***$P < 0.001$). Results show mean and s.e.m.

features of these functionally distinct populations, we reconstructed projection-identified BLA neurons. We observed greater distal dendritic branching in CeM projectors (Fig. 4h, i), though NAc projectors and CeM projectors contained both pyramidal and stellate cells (Fig. 4i inset and Extended Data Fig. 8).

Finally, we compared the transcriptomes of BLA neurons projecting to the NAc or CeM (Fig. 4j and Extended Data Fig. 9). Following retrobead injections into the NAc or CeM, we dissociated labelled BLA neurons and performed RNA-seq (Fig. 4j). RNA-seq revealed relatively few candidate genes expressed differentially between NAc and CeM projectors, consistent with the idea that these two populations are closely related (Fig. 4k and Extended Data Fig. 9). However, these differentially expressed candidate genes may underpin the mechanisms that contribute to the wiring of these distinct populations through development and/or rapidly biasing gain modulation of synaptic transmission during valence-specific learning.

Taken together, NAc and CeM projectors are populations of BLA neurons that undergo opposing synaptic changes following fear or reward conditioning, and optogenetic manipulation of NAc and CeM projectors reveals causal relationships with valence-specific behaviours. Further, we have identified distinguishing electrophysiological, morphological and gene expression characteristics that facilitate further investigation. Our study suggests that the indelible nature of valence encoding observed in amygdala neurons[10] is mediated by connectivity, and the topographical intermingling of these populations may serve to facilitate interaction[30]. In conclusion, the BLA is a site of divergence for circuits mediating positive and negative emotional or motivational valence.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Tye, K. M., Stuber, G. D., De Ridder, B., Bonci, A. & Janak, P. H. Rapid strengthening of thalamo-amygdala synapses mediates cue–reward learning. *Nature* **453,** 1253–1257 (2008).
2. Tye, K. M. *et al.* Methylphenidate facilitates learning-induced amygdala plasticity. *Nature Neurosci.* **13,** 475–481 (2010).
3. McKernan, M. G. & Shinnick-Gallagher, P. Fear conditioning induces a lasting potentiation of synaptic currents *in vitro. Nature* **390,** 607–611 (1997).
4. Rogan, M. T., Stäubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390,** 604–607 (1997).
5. Rumpel, S., LeDoux, J., Zador, A. & Malinow, R. Postsynaptic Receptor Trafficking Underlying a Form of Associative Learning. *Science* **308,** 83–88 (2005).
6. Maren, S. Synaptic mechanisms of associative memory in the amygdala. *Neuron* **47,** 783–786 (2005).
7. Han, J.-H. *et al.* Selective erasure of a fear memory. *Science* **323,** 1492–1496 (2009).
8. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439,** 865–870 (2006).
9. Shabel, S. J. & Janak, P. H. Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proc. Natl Acad. Sci. USA* **106,** 15031–15036 (2009).
10. Redondo, R. L. *et al.* Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature* **513,** 426–430 (2014).
11. Sah, P., Faber, E. S. L., Lopez De Armentia, M. & Power, J. The amygdaloid complex: anatomy and physiology. *Physiol. Rev.* **83,** 803–834 (2003).
12. Romanski, L. M., Clugnet, M.-C., Bordi, F. & LeDoux, J. E. Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behav. Neurosci.* **107,** 444–450 (1993).
13. Fontanini, A., Grossman, S. E., Figueroa, J. A. & Katz, D. B. Distinct subtypes of basolateral amygdala taste neurons reflect palatability and reward. *J. Neurosci.* **29,** 2486–2495 (2009).
14. Pitkänen, A. in *The Amygdala: A Functional Analysis* 2nd edn (ed. Aggleton, J. P.) Ch. 2 (Oxford Univ. Press, 2000).
15. Davis, M. The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* **15,** 353–375 (1992).
16. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26,** 321–352 (2002).
17. Uwano, T., Nishijo, H., Ono, T. & Tamura, R. Neuronal responsiveness to various sensory stimuli, and associative learning in the rat amygdala. *Neuroscience* **68,** 339–361 (1995).
18. Clem, R. L. & Huganir, R. L. Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. *Science* **330,** 1108–1112 (2010).
19. Miserendino, M. J. D., Sananes, C. B., Melia, K. R. & Davis, M. Blocking of acquisition but not expression of conditioned fear-potentiated startle by NMDA antagonists in the amygdala. *Nature* **345,** 716–718 (1990).
20. Fanselow, M. S. & LeDoux, J. E. and others. Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron* **23,** 229–232 (1999).
21. Stuber, G. D. *et al.* Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* **475,** 377–380 (2011).
22. Ambroggi, F., Ishikawa, A., Fields, H. L. & Nicola, S. M. Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* **59,** 648–661 (2008).
23. Jimenez, S. A. & Maren, S. Nuclear disconnection within the amygdala reveals a direct pathway to fear. *Learn. Mem.* **16,** 766–768 (2009).
24. Haubensak, W. *et al.* Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature* **468,** 270–276 (2010).
25. Ciocchi, S. *et al.* Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468,** 277–282 (2010).
26. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53,** 639–647 (2007).
27. Paré, D., Smith, Y. & Paré, J.-F. Intra-amygdaloid projections of the basolateral and basomedial nuclei in the cat: *Phaseolus vulgaris*-leucoagglutinin anterograde tracing at the light and electron microscopic level. *Neuroscience* **69,** 567–583 (1995).
28. Tye, K. M. *et al.* Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471,** 358–362 (2011).
29. Washburn, M. S. & Moises, H. C. Electrophysiological and morphological properties of rat basolateral amygdaloid neurons *in vitro. J. Neurosci.* **12,** 4066–4079 (1992).
30. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. *Nature* **517,** 284–292 (2015).

METHODS

Animals and stereotaxic surgery. Adult wild-type male C57BL/6 mice (248 mice), aged 6–12 weeks (8.3 ± 1.5 weeks; Jackson Laboratory or Charles River Laboratories for RNA-seq) were used for experiments. Following surgery, animals were maintained on a reverse 12 h light/dark cycle with *ad libitum* food and water. All procedures of handling animals were in accordance with the guidelines from the NIH, and with approval of the MIT or Harvard Medical School Institutional Animal Care and Use Committee. All surgeries were conducted under aseptic conditions using a digital small animal stereotaxic instrument (David Kopf Instruments).

Mice were anaesthetized with isoflurane (5% for induction, 1.5–2.0% afterward) in the stereotaxic frame for the entire surgery and their body temperature was maintained with a heating pad. In order to label basolateral amygdala (BLA) neurons projecting to the nucleus accumbens (NAc), about 70 nl of red or green retrobeads (RetroBeads, Lumafluor Inc.) were injected into the NAc at stereotaxic coordinates from bregma: +1.4 mm anteroposterior, ±0.87 mm mediolateral and −4.7 mm dorsoventral. In order to label BLA neurons projecting to the medial part of the central amygdala (CeM), 50 nl of retrobeads (different colour from NAc injection) were injected in the contralateral CeM (−0.75 mm anteroposterior, ±2.35 mm mediolateral and −5.08 mm dorsoventral). To test the causal role of BLA neurons projecting to NAc or to CeM in reward and aversive behaviours, we injected a retrogradely travelling rabies virus carrying channelrhodopsin2–Venus fusion protein (RV-4ChopV(B19G)) or Venus[31,32] (Fig. 2 and Extended Data Fig. 4) into NAc (250 nl) or CeM (180 nl). These virus constructs are referred to as RV-ChR2–Venus and RV-Venus in the manuscript. All animals receiving control virus injection were pooled into one control group. Injections were performed using glass micropipettes (1–5 μl; Drummond Scientific) pulled with a puller (Narishige PC-10) mounted on 10-μl microsyringes (Hamilton Microlitre 701; Hamilton Co.) to deliver the retrobeads at a rate of 2 nl s^{-1} or virus at a rate of 0.5 to 1 nl s^{-1}, using a microsyringe pump (UMP3; WPI) and controller (Micro4; WPI). After completion of the injection, the pipette was raised 100 μm and left for an additional 10 min to allow diffusion of the retrobeads or the virus at the injection site and then slowly withdrawn.

In a separate group of experiments (Fig. 3), adeno-associated virus serotype 5 (AAV$_5$) carrying halorhodopsin 3.0 fused to enhanced yellow fluorescent protein (eYFP) in a double-floxed inverted open reading frame (DIO) under the control of EF1α promoter (AAV$_5$-EF1α-DIO-eNpHR3.0-eYFP) or a corresponding fluorophore control (AAV$_5$-EF1α-DIO-eYFP) was injected bilaterally into the BLA (400 nl in each hemisphere) at stereotaxic coordinates from bregma: −1.60 mm anteroposterior, ±3.35 mm mediolateral and −4.90 mm dorsoventral. Concurrently, canine adenovirus 2 (CAV2)[33,34] carrying Cre-recombinase (or a mixture of CAV2-Cre and herpes simplex virus carrying Cre-recombinase and the fluorescent reporter mCherry under the control of EF1α promoter) was injected into the NAc (300 nl in each hemisphere) or CeM (150 nl in each hemisphere) at the same coordinates described above. We summarize this strategy with the designation 'Cre-DIO'. All animals receiving fluorophore control injection were pooled into one control group. In this manuscript, CAV refers to CAV2, and NpHR refers to eNpHR3.0.

In order to deliver light to the BLA, a 300-μm diameter optic fibre (0.37 numerical aperture (NA)) glued to a 1.25 mm ferrule was implanted above the BLA (−1.64 mm anteroposterior, ±3.35 mm mediolateral and −4.25 mm dorsoventral). One layer of adhesive cement (C&B metabond; Parkell) followed by cranioplastic cement (Dental cement; Stoelting) was used to secure the fibre ferrule to the skull and, 20 min later, the incision was sutured. Blue light (~20 mW, ~283 mW per mm^2 at the fibre tip) was delivered using a 473 nm laser and yellow light (~10 mW, ~141 mW per mm^2 at the fibre tip) was delivered using a 589/593 nm laser.

After surgery, body temperature was maintained using a heat lamp until the animal fully recovered from anaesthesia. Behavioural experiments followed by *ex vivo* electrophysiological recordings were conducted approximately 2 weeks after surgery. Behavioural experiments requiring expression via rabies virus were conducted 4–5 days after surgery (Fig. 2). Behavioural experiments requiring viral expression via the Cre-DIO strategy were conducted 3 months after surgery (Fig. 3).

Pavlovian conditioning. *Fear conditioning.* The mice were conditioned using behavioural hardware boxes (MedAssociate) placed in custom-made sound-attenuating chambers. Each box contained a modular test cage with an electrifiable grid floor and a speaker. On the day of conditioning, all animals were exposed to two 30 min conditioning sessions separated by 20 min in the home cage (Extended Data Fig. 1e). During the first session, mice in the unpaired group received six tone presentations, while mice in the paired group received no cues. During the second session, mice in the unpaired group received six foot shocks whereas mice in the paired group received six tones, each co-terminating with a shock (paired fear). During each session, a period of acclimation lasting 200 s preceded the presentation

of the first tone or foot shock. The conditioned stimulus consisted of a 2 kHz, 80 dB pure tone lasting 20 s. The unconditioned stimulus consisted of scrambled 1.5 mA foot shock lasting 2 s. During paired conditioning, the conditioned and unconditioned stimuli were co-terminating. Cue presentations were separated by 70 to 130 s. Following conditioning, mice were returned to their home cages for ~24 h until preparation of brain slices (Extended Data Fig. 1e). Videos of the mice were acquired during the second session for the paired group and both sessions for the unpaired group, to allow freezing quantification during the tone (an infrared LED was switched on during the period of the tone to synchronize CS with the video). *Measurement of fear behaviour.* Percentage time freezing during conditioned stimulus presentation (Extended Data Fig. 1f) was quantified by manual scoring with the help of custom-written software in MATLAB. For each of the six trials, a segment of the video containing 20 s of the tone and the 20 s preceding the tone was extracted frame by frame and exported into MATLAB. Two additional segments of the same duration were extracted starting at randomly chosen points in the video. The sequence of these eight trials was then randomized and presented to a human scorer who was blind to the trial number. The scores were then re-assembled and the percentage of freezing during the conditioned stimulus was calculated. Frame numbers of the video containing the tone were confirmed by generating a heat map of the intensity of the pixels containing the infrared LED. *Reward conditioning.* Twenty hours after food restriction, the mice were conditioned in sound-attenuating boxes (MedAssociates). Each box contained a modular test cage assembled with a sucrose delivery port, a speaker and a house light placed under the sucrose port. The conditioned stimulus consisted of a compound light–tone cue, ended by a beam break, 400 ms after port entry detection. If the mouse did not enter the port, the tone lasted for 30 s. The tone was a 5 kHz, 80 dB pure tone. For the paired group, a small volume (15 μl) of a sucrose solution (30%) was delivered into the port 1 s after conditioned stimulus onset, only if the mouse had entered the port after the onset of the previous conditioned stimulus, to prevent sucrose accumulation. For the unpaired group, no sucrose was delivered to the port. The inter-trial interval (ITI) of the conditioned stimulus presentations was chosen randomly from a list at runtime and was 143 ± 40 s for the first 20 conditioned stimulus presentations and 108 ± 32 s for the 100 subsequent conditioned stimulus presentations. The conditioning session was terminated after 120 sucrose deliveries and lasted for about 4 h. Performance of the mouse was assessed during the second half of a conditioning session. After the training session, the unpaired mice received the same amount of sucrose as the paired mouse (in their home cage 15 μl × 120 = 1.8 ml, Extended Data Fig. 1e). In the paired group, if the mice did not claim all the sucrose, the remaining volume was made available to the mouse in its home cage after the conditioning session. After 20 min in their home cages, all the animals had *ad libitum* food until the electrophysiology experiment the following day.

Learning criterion for reward learning. A mouse was considered to have acquired the task (and classified as a learner) if the number of port entries from +1 s to +8 s relative to conditioned stimulus onset (Extended Data Fig. 1g, black line) was significantly higher than the number of port entries from −1 s to −8 s relative to conditioned stimulus onset (Extended Data Fig. 1g, grey line). Statistical significance was tested using a one-sided Wilcoxon rank-sum test (MATLAB) and the threshold for learning was set at $P < 0.001$.

Conditioning for photoinhibition experiments. In these experiments, the same animals experienced reward conditioning followed by fear conditioning after 1–4 weeks. Pure tones of 2 and 10 kHz were used as conditioning stimuli in these two paradigms; the use of these two tones was counterbalanced across animals. Animals were tethered using a dual commutator for light delivery. During fear conditioning, yellow light was delivered 1 s before shock onset until 1 s after shock termination (shock intensity = 0.5 mA). During reward learning, yellow light was delivered for 7.5 s immediately following a port entry during conditioned stimulus (tone) presentation.

Ex vivo electrophysiology. *Brain tissue preparation.* About 2 weeks after retrobead injections in NAc and CeM, 114 mice were anaesthetized with 90 mg kg^{-1} pentobarbital sodium and perfused transcardially with 10 ml of modified artificial cerebrospinal fluid (ACSF) at ~4 °C saturated with 95% O$_2$ and 5% CO$_2$, containing: 75 mM sucrose, 87 mM NaCl, 2.5 mM KCl, 1.3 mM NaH$_2$PO$_4$, 7 mM MgCl$_2$, 0.5 mM CaCl$_2$, 25 mM NaHCO$_3$ and 5 mM ascorbic acid (pH 7.25–7.4, 327 ± 3 mOsm). The brain was then extracted and glued (Roti coll 1; Carh Roth GmbH) on the platform of a semiautomatic vibrating blade microtome (VT1200; Leica). The platform was then placed in the slicing chamber containing modified ACSF at 4 °C. Coronal sections of 300 μm containing the NAc, CeM or BLA were collected in a holding chamber filled with ACSF saturated with 95% O$_2$ and 5% CO$_2$, containing: 126 mM NaCl, 2.5 mM KCl, 1.25 mM NaH$_2$PO$_4$, 1.0 mM MgCl$_2$, 2.4 mM CaCl$_2$, 26.0 mM NaHCO$_3$, 10 mM glucose (pH 7.25–7.4, 298 ± 2 mOsm). Recordings were started 1 h after slicing and the temperature was maintained at approximately 31 °C both in the holding chamber and during the recordings.

All retrobead injection sites were checked and imaged with a camera (Hamatsu) attached to the microscope (BX51; Olympus, Extended Data Fig. 1). The slice images were reported on the mouse brain atlas (Paxinos and Watson) and the centre of the injection was taken at the brightest point of the fluorescence. Some of the retrobead injections had dorsoventral leaks. In this case, the centre of the injection was taken on the brightest fluorescent point in the target structure (Extended Data Fig. 1). If the injection site was outside NAc or CeM, respective projectors of this injection were not recorded in the BLA. In addition, all the CeM injections were overlaid on the mouse atlas. CeM projector recordings collected from animals with injection sites that had leaks into CeC and/or CeL were discarded.

Whole-cell patch-clamp recording. Recordings were made from visually identified neurons containing retrobeads. Patched cells were filled with Alexa Fluor 350 and biocytin, visualized and superimposed with retrobead fluorescence to confirm whether the patched cell was retrobead positive.

For measuring AMPAR/NMDAR ratio, brain slices containing the BLA were then placed in the recording chamber perfused with ACSF containing 100 μM of the γ-aminobutyric acid (GABA$_A$R) receptor antagonist picrotoxin (R&D systems). Picrotoxin was not used to assay passive membrane properties from NAc- and CeM-projecting BLA neurons. A bipolar stimulating electrode (~80 μm spacing between tips) was placed in the amygdala–striatal transition zone containing internal capsule fibres (Extended Data Fig. 2). Electric stimulation intensity was between 0.01 and 0.2 mA. For electrophysiological characterization of NAc and CeM projectors (Fig. 3a–g) and confirmation of the health of rabies-virus-transduced cells (Extended Data Fig. 7), picrotoxin was not added to the ACSF.

Voltage-clamp recordings were made using glass microelectrodes (4–6 MΩ) shaped with a horizontal puller (P-1000) and filled with a solution containing: 120 mM caesium methansulphonate, 20 mM HEPES, 0.4 mM EGTA, 2.8 mM NaCl, 5 mM tetraethylammonium chloride, 2.5 mM MgATP, 0.25 mM NaGTP, 8 mM biocytin and 2 mM Alexa Fluor 350 (pH 7.3, 283 mOsm). The cells were first clamped at −70mV to determine optimal intensity for the electric stimulation of the internal capsule. Current-clamp recordings to characterize electrophysiological properties of NAc and CeM projectors were made using similar glass microelectrodes (4–6 MΩ) filled with a solution containing: 125 mM potassium gluconate, 20 mM HEPES, 10 mM NaCl, 3 mM MgATP, 8 mM biocytin and 2 mM Alexa Fluor 350 (pH 7.3; 283 mOsm).

Recorded signals were amplified using Multiclamp 700B amplifier (Molecular Devices). Analogue signals were digitized at 10 kHz using a Digidata 1440 and pClamp9 software (Molecular Devices). ACSF and drugs were applied to the slice via a peristaltic pump (Minipuls3; Gilson) at 3 ml per min.

All recordings were performed blind to the performance of the animal, and a subset of the data was obtained blind to the behavioural conditioning group of the animal. We recorded the position of the cells within the BLA (Extended Data Fig. 2) and the placement of the stimulating electrode relative to the BLA. There was no observable difference in either of these parameters across slices obtained from animals in different behavioural conditioning groups.

In order to obtain the AMPAR/NMDAR ratio, the cell was first voltage clamped at +40 mV. Once we obtained a stable baseline excitatory post-synaptic current (EPSC) amplitude in response to internal capsule fibre stimulation (compound AMPAR + NMDAR current), we bath applied the NMDAR antagonist AP5 (D-(-)-2amino-5-phosphonopentanoate; R&D systems) at a concentration of 50 μM. AMPAR EPSCs were recorded starting from 5 min after the action of AP5. NMDAR current was obtained by subtracting the average EPSC trace of the AMPAR current from the compound current. Each group in Fig. 1 included 9–13 neurons recorded from 6–9 mice.

Histology. The location of all recorded neurons was checked after the recording. Co-localization of Alexa Fluor 350 and retrobeads was confirmed at the end of the recording, and double-checked with confocal microscopy for the cells that were recovered with streptavidin staining. For each experiment, the slices containing a retrobead injection site (NAc and CeM) or a recorded neuron (BLA) were fixed overnight at 4 °C in 4% paraformaldehyde (PFA), and then kept in phosphate buffered saline (PBS). Slices containing patched neurons were incubated for 2 h in streptavidin-CF405 (2 mg ml^{-1}, dilution 1:500; Biotium), mounted on microscope slides with PVA-DABCO and imaged under the confocal microscope (Extended Data Fig. 3).

Data analysis. Offline analysis of AMPAR/NMDAR ratios, paired-pulse ratios and electric properties was performed using Clampfit (Molecular Devices) and MATLAB software written by P.N. Membrane properties including access resistance of the cell were computed by a MATLAB implementation of the Q-method[35]. All custom-written software is available upon request.

In vivo optogenetic behaviour. *Light delivery.* For optical stimulation during behavioural assays (Fig. 2), a 473 nm (blue light) or 589/593 nm laser (yellow light) (OEM Laser Systems) was connected to a patch cord with a pair of FC/PC

connectors in each end (Doric). This patch cord was connected through a fibre optic rotary joint (which allows free rotation of the fibre; Doric) with another patch cord with an FC/PC connector on one side and a ferrule connection on the other side (matching the size of the ferrule glued to the optic fibre implanted in the mouse). The optic fibre implanted in the mouse (300 μm diameter, 0.37 NA) was connected to the optic patch cord using ceramic mating sleeves (PFP). Blue light was delivered at 20 mW in 20 Hz, 5 ms light pulses. Yellow light was delivered at 10 mW, for 4 s during fear conditioning and 7.5 s during reward conditioning. Laser output was modulated with a Master 8 pulse stimulator (A.M.P.I.). Onset of laser pulses was determined by behavioural hardware (MedPC Associates).

Intracranial optical self-stimulation (ICSS). 4.5 days after surgery, mice were food-restricted overnight. Immediately before the start of each session, fibre optic implants were connected to a patch cord and the mouse was placed in conditioning chambers equipped with active and inactive nose-poke ports directly below two cue lights, as well as auditory stimulus generators and video cameras. Mice were given two self-stimulation sessions on two consecutive days (5.5 and 6.5 days after surgery) in which they could respond freely at either nose-poke port. On day 1 both nose-poke ports were baited with a crushed cereal treat to facilitate initial investigation. The start of each 2 h session was indicated by the illumination of both nose-poke ports and the onset of low-volume white noise to mask unrelated sounds. Each nose poke in the active port resulted in light stimulation of BLA projectors (60 pulses, 20 Hz, 5 ms pulse duration). Concurrently, the cue-light above the respective port was illuminated and a distinct 1-s tone was played for each nose poke (1 kHz and 1.5 kHz counterbalanced), providing a visual and auditory cue whenever a nose poke occurred. Active and inactive nose-poke time stamps were recorded using Med-PC software and day 2 data were analysed using MATLAB and Microsoft Excel.

Real-time place avoidance (RTPA). The RTPA chamber was constructed from transparent plastic (50 × 53 cm) and divided into two equal compartments. One of these was assigned as the photo-stimulated zone (counterbalanced between animals). At the start of the 1 h session, individual mice were placed in the non-stimulated side of the chamber. Every time the mouse crossed to the side of the chamber paired with photostimulation, 20 Hz (~20 mW, 5 ms pulse duration) laser stimulation was delivered until the mouse crossed back into the non-stimulated side. Ethovision XT video tracking software (Noldus Information Technologies) was used to track the animal and control the onset and offset of light pulse trains. Data were subsequently analysed using MATLAB and Microsoft Excel software.

Histology. After optogenetic experiments, all mice were anaesthetized with pentobarbital sodium, and transcardially perfused with ice-cold Ringer's solution followed by ice-cold 4% PFA in PBS (pH 7.3). Extracted brains were fixed in 4% PFA overnight and then equilibrated in 30% sucrose in PBS. 40-μm thick coronal sections were sliced using a sliding microtome (HM430; Thermo Fisher Scientific) and stored in PBS at 4 °C until they were processed for histology. Sections were then incubated with a DNA-specific fluorescent probe (DAPI: 4′,6-diamidino-2-phenylindole (1:50,000)) for 30 min, and finally washed with 1× PBS followed by mounting on microscope slides with PVA-DABCO.

Imaging. Confocal fluorescence images were acquired on an Olympus FV1000 confocal laser scanning microscope using a 10×/0.40 NA objective for viral injections and fibre placements imaging or a 40×/1.30 NA or 60×/1.42 NA oil-immersion objectives for imaging streptavidin-CF405 stained neurons. The centre of the viral injection was taken at the brightest fluorescent point in anteroposterior, mediolateral and dorsoventral axis. The tip of the fibre was determined by the ~50-μm thick gliosis generated by the fibre. Neurons recovered from the streptavidin staining (~60% recorded in the whole cell) were imaged covering the whole dendritic and axonal arbourization contained in the slice.

Neuron reconstructions. Imaris software (Bitplane Inc.) was used to reconstruct neurons from z-stacks of confocal images and to perform Sholl analysis[36]. Since we are reconstructing neurons filled during whole-cell patch clamp recordings, we chose to focus solely on dendritic branching patterns and did not examine parameters such as the volume of the soma, since this measure might be compromised by experimental procedures. Overlaying the atlas on images of brain slices revealed that the distances depicted in the atlas are about 90% smaller than physical distances in brain slices. For example, if the length of a brain region is 900 μm in the atlas, then the real distance measured in our brain sections was about 1000 μm. Therefore, to depict the reconstructed neurons (Fig. 4h, i and Extended Data Fig. 8), we multiplied physical distances by a factor of 0.9.

RNA sequencing (RNA-seq). *Manual cell sorting and RNA sequencing.* The RNA-seq experiment was repeated twice to verify reproducibility, and we refer to these biological replicates as experiment 1 and experiment 2 (see Extended Data Fig. 9c: samples used in experiment 1 and 2 are indicated in black ($n = 9$) and blue ($n = 8$) respectively below the heatmap). Manual sorting of fluorescent cells were carried out as described in ref. 37. In brief, adult C57Bl6 male

mice (~6–8-weeks-old at surgery, C57BL/6NCrl for experiment 1, C57BL/6J for experiment 2) were injected with red fluorescent retrograde beads into either NAc or CeM. Surgeries for experiment 1 were conducted at Harvard Medical School by P.N. and surgeries for experiment 2 were conducted at MIT by G.G.C. About 2 weeks after surgery, animals were decapitated under isoflurane anaesthesia and their brain was quickly removed and transferred into ice-cold oxygenated ACSF, containing 126 mM NaCl, 20 mM NaHCO$_3$, 20 mM dextrose, 3 mM KCl, 1.25 mM NaH$_2$PO$_4$, 2 mM CaCl$_2$, 2 mM MgCl$_2$, 50 µM AP5, 20 µM DNQX and 100 nM TTX. Acute 330-µm coronal brain slices were prepared and incubated in a protease (1.2 mg ml^{-1} protease E; Sigma-Aldrich) containing oxygenated ACSF for 50 min. After 15 min of washes in the ACSF, BLA tissue was microdissected using a pair of fine scissors under a fluorescent dissecting microscope (Leica M165FC stereomicroscope). The dissected BLA tissue was then triturated in ACSF using a series of three Pasteur pipettes of decreasing tip diameters and the dissociated cells were transferred into a small Petri dish. With visual control under a fluorescent dissecting microscope, red-retrobead-positive neurons were aspirated into a micropipette with a 30–50 µm tip diameter and transferred into a clean Petri dish. A total of 35–60 retrobead-positive neurons were pooled for each sample, which were immediately lysed in 50 µl of extraction buffer (PicoPure RNA isolation kit, Arcturus, Life Technologies) and total mRNA was subsequently isolated. Complementary DNA was synthesized using Ovation RNA-seq System V2 kit (Neugen). We obtained approximately 6 µg of cDNA from 35–60 cells from each group. Then, the cDNA library was prepared using Ovation Ultralow DRMultiplex Systems (Nugen). Sequencing was conducted on an Illumina HiSeq2500 using single-end 50 base pairs at the Biopolymer facility, Harvard Medical School (for experiment 1), and an illumina NextSeq 500 using single-end 75 base pairs with high flow cell size at the FAS centre for systems biology, Harvard University (for experiment 2). The total number of reads that we obtained for each sample were approximately 34 million (for experiment 1) and 50 million (for experiment 2).

Analysis of RNA-seq data. Sequencing reads were mapped using Tophat version 2.0.10 (http://ccb.jhu.edu/software/tophat/index.shtml) against the *Mus musculus* UCSC version mm10 genome. After alignment, the read counts for each gene were extracted using HTseq (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html) based on the mm10 Refseq gff file. Log2 fold differences were computed from each of two independent experiments using DEseq2. Candidate differentially expressed genes were required to be enriched in CeM or NAc projectors at a quantile fold-difference threshold of 0.01 (Fig. 4) or 0.02 (Extended Data Fig. 9) in each of two independent experiments ($n = 8$, NAc; $n = 9$, CeM total). To estimate false discovery rate (Extended Data Fig. 9), we used two types of chance estimates. One of the chance estimates, 'flip-flopped', is taken from genes that passed the quantile thresholds but were enriched in opposite populations in the two experiments. Another chance estimate, 'permuted', is determined based on permuting fold differences across genes within each independent experiment (Extended Data Fig. 9). We estimated false discovery rate using the more conservative chance estimate, flip-flopped with the following formula: (number of genes on the flip-flopped list/number of genes on the differentially expressed genes). RNA-seq data has been deposited in the Gene Expression Omnibus under accession code GSE66345

Statistical analysis. Statistical analyses were performed using commercial software (GraphPad Prism; GraphPad Software, Inc.). Within-subject comparisons were made using paired tests. Group differences were detected using either one-way analysis of variance (ANOVA) or with two-way ANOVA, both followed by Bonferroni post-hoc tests. Corrections for multiple comparisons were made when appropriate. The reported numbers of degrees of freedom (df) for each one-way ANOVA are between column degrees of freedom and total degrees of freedom. Since normality tests have little power to detect non-gaussian distributions with small data sets, we did not explicitly test for the normality of our data sets. We used the Grubbs' test to detect and remove outliers from our data. Single-variable differences were detected with two-tailed paired or unpaired (as noted) Student's *t*-tests. For all results, the significance threshold was placed at $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$), and corrections for multiple comparisons were reflected in the P value rather than in α. All data are shown as mean and s.e.m. To assess learning during the reward task, we used a one-sided Wilcoxon rank-sum test (MATLAB) and set the threshold for learning at $P < 0.001$. Result sheets of statistical tests from GraphPad software detailing (wherever applicable) estimates of variance within each group, confidence intervals, effectiveness of pairing (in case of paired *t*-tests), comparison of variances across groups, etc. are available upon request.

Sample size. The target number of samples in each group was determined based on numbers reported in published studies. No statistical methods were used to predetermine sample size. In the photoinhibition experiment, since the viral incubation time was long, we factored in the skill of the surgeon to determine the number of surgeries to be performed. Our target number of animals in each group was 12. The experimenter performing surgeries was known to hit the targets used (NAc, CeM, BLA) with a probability of 0.9. Since there were six targets in each brain (three in each hemisphere—two injections and one optical fibre), the probability of a successful surgery would be approximately $(0.9)^6 \approx 0.5$. We therefore performed about 24 surgeries in each group. All sample sizes mentioned in figures represent biological replicates. All animals receiving control virus injections were pooled into one control group (Figs 2 and 3). In the Venus group from Fig. 2c, there were six animals with injections in NAc and six animals with injections in CeM. In the Venus group from Fig. 2e, there were four animals with injection in NAc and two animals with injection in CeM. In the eYFP group from Fig. 3b, there were ten animals with injections in NAc and six animals with injections in CeM. In the eYFP group from Fig. 3c, there were nine animals with injections in NAc and five animals with injections in CeM.

Replication. Results from AMPAR/NMDAR ratio experiments were replicated once with a different experimenter and the final numbers reported in the paper are pooled across both repetitions of the experiment (Fig. 1). Photostimulation and photoinhibition experiments (Figs 2 and 3) were not replicated. The RNA-seq experiment was also replicated once (Fig. 4).

Randomization. All surgical and behavioural manipulations performed on each animal were determined randomly. All randomization was performed by an experimenter, and no explicit randomization algorithm was used. For animals used in photostimulation and photoinhibition experiments, the virus used in each animal (ChR2/Venus or NpHR/eYFP) and injection site (NAc/CeM) were determined randomly and the stereotaxic apparatus used for surgery was counterbalanced across groups. For surgeries with unilateral injections and/or fibre placements, the hemisphere used for injections was determined randomly during the time of the surgery. Surgeries were performed on animals caged in groups of 4 or 5 animals. Animals from each cage were allocated to at least two behavioural groups. All animals used in *ex vivo* electrophysiology experiments were isolated at least 1 day before behavioural conditioning.

Exclusion criteria. *Ex-vivo electrophysiology.* Data were excluded based on predetermined histological and electrophysiological criteria, established during pilot experiments. The injection site was determined as the most ventral point where fluorescence was brightest, and data from cells where the corresponding retrobead injection was outside the target region (NAc or CeM) were excluded (Extended Data Fig. 1a–d). Each recorded cell was confirmed to be a projector by overlaying the fluorescence from retrobeads with the fluorescence from Alexa Fluor dye contained in the pipette. Each cell was also confirmed to be in the BLA by visualizing with differential interference contrast microscopy under a 4× objective. We did a secondary confirmation under the confocal microscope for cells that were recovered from streptavidin staining. Cells in which evoked responses were polysynaptic (multiple peaks in the evoked current) were discarded. Data from cells whose access resistance was greater than 40 MΩ or cells that died during recording were also excluded.

Animals used in photostimulation and photoinhibition experiments. Data from animals used in photostimulation and photoinhibition experiments (Figs 2 and 3) were excluded based on histological and performance criteria established during pilot experiments. Histological criteria included injection sites and optical fibre placement (Extended Data Figs 4 and 5). Only animals with injection sites in the region of interest (NAc or CeM) were included. For animals with a rabies virus injection in CeM, atlas outlines were overlaid manually over a confocal image of the BLA containing the damage caused by the tip of the optic fibre. Light cones based on numerical aperture of the optic fibre (NA 0.37, ~15° half angle) were then drawn below the optic fibre and animals in which light cones encompassed central amygdala were excluded from further analysis. For the optical self-stimulation experiment, data from animals that did not respond at least 40 times (sum of nose pokes in active and inactive ports) over the 2 hour period were excluded from further analysis. For photoinhibition experiments, the amount of expression in each hemisphere of the BLA was rated on a scale of 0–5 based on fluorescence intensity by an experimenter blind to the behavioural performance of the animal. These ratings were gathered in an excel sheet, read by a MATLAB script and only data from animals with fluorescence ratings greater than 4 in each hemisphere were included for further analysis.

31. Kiritani, T., Wickersham, I. R., Seung, H. S. & Shepherd, G. M. G. Hierarchical connectivity and connection-specific dynamics in the corticospinal–corticostriatal microcircuit in mouse motor cortex. *J. Neurosci.* **32,** 4992–5001 (2012).
32. Wickersham, I. R., Sullivan, H. A. & Seung, H. S. Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nature Protocols* **5,** 595–606 (2010).
33. Soudais, C., Boutin, S. & Kremer, E. J. Characterization of cis-Acting Sequences Involved in Canine Adenovirus Packaging. *Mol. Ther.* **3,** 631–640 (2001).

34. Hnasko, T. S. *et al.* Cre recombinase-mediated restoration of nigrostriatal dopamine in dopamine-deficient mice reverses hypophagia and bradykinesia. *Proc. Natl Acad. Sci. USA* **103,** 8858–8863 (2006).

35. Novák, P. & Zahradník, I. Q-method for high-resolution, whole-cell patch-clamp impedance measurements using square wave stimulation. *Ann. Biomed. Eng.* **34,** 1201–1212 (2006).

36. Sholl, D. A. Dendritic organization in the neurons of the visual and motor cortices of the cat. *J. Anat.* **87,** 387–406 (1953).

37. Hempel, C. M., Sugino, K. & Nelson, S. B. A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nature Protocols* **2,** 2924–2929 (2007).
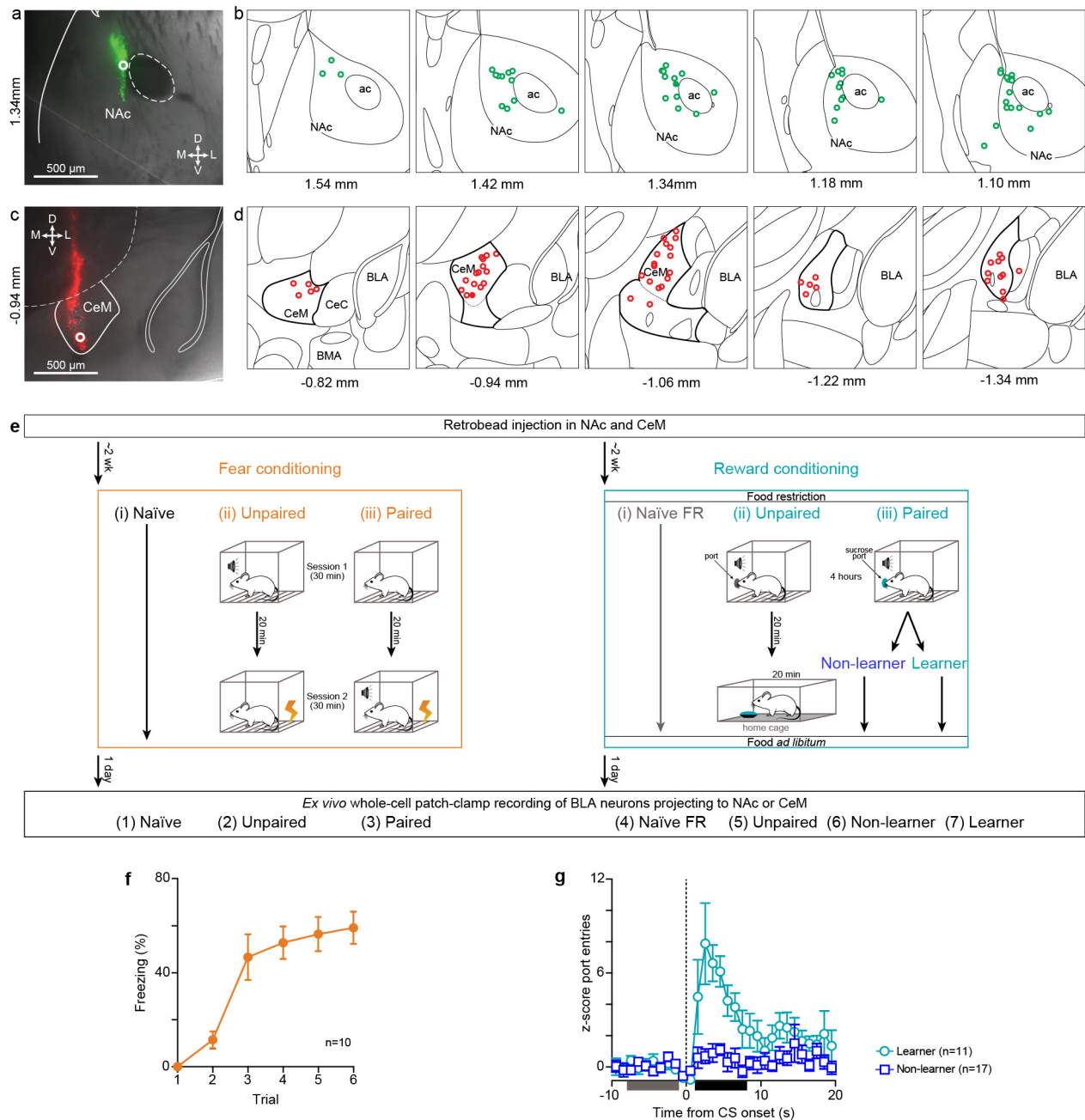
Extended Data Figure 1 | Histological verification of retrobead injection sites and behavioural quantification of fear and reward conditioning for mice used in Fig. 1. a, Representative differential interference contrast (DIC) image of a 300-μm thick coronal slice containing the centre of the retrobead injection in NAc . The white circle indicates the most ventral point at which fluorescence is brightest and corresponds to the filled green circle in **b**. **b**, Location of all retrobead injection sites (green circles) in the NAc for all mice used in Fig. 1. Each atlas schematic represents a 1.5 mm × 1.5 mm region of the atlas and the corresponding anteroposterior stereotaxic coordinate relative to Bregma is indicated below. **c**, Representative DIC image of a 300-μm thick coronal slice containing the centre of the retrobead injection in CeM as indicated by the white dot. **d**, Retrobead injection sites in CeM (red circles) for all mice used in Fig. 1, with the example from **c** indicated by the filled red circle. The corresponding anteroposterior stereotaxic coordinate relative to Bregma is indicated below. **e**, Experimental design for AMPAR/NMDAR ratios from Fig. 1. Either red or green retrobeads were injected in the NAc and the other colour in the contralateral CeM. Two weeks after injection, the retrobeads had travelled back to the cell bodies of the BLA neurons projecting to NAc or CeM. Animals were conditioned 1 day before *ex vivo* whole-cell patch-clamp recordings. Each mouse received one of six conditioning protocols, three protocols categorized under 'fear conditioning' and three protocols categorized under 'reward conditioning'. Fear conditioning protocols: (i) naive, animals were naive to the operant chamber. (ii) Unpaired, animals were exposed to the conditioning chamber in two sessions. Animals received six tones in the first session and they received six foot shocks in the second session. Animals were returned to their home cage for ~20 min between the two sessions. (iii) Paired, animals were exposed to the operant chamber in two sessions. Animals did not receive any tone or shock stimuli in the first session, and received tones co-terminating with shocks in the second session. Animals were returned to their home cage for 20 min between the two sessions. Protocols for unpaired and paired fear groups were adapted from ref. 18. Reward conditioning protocols: (i) naive food restricted (FR), animals naive to the operant chamber were food restricted two days before *ex vivo* experiments and had free access to food for 1 day before *ex vivo* experiments. We used this group to control for changes in synaptic strength caused by food restriction which was necessary in reward conditioning groups to expedite task acquisition, adapted from rats as in refs 1 and 2. (ii) Unpaired, animals received tones in the operant chamber, were returned to their home cage for ~20 min after which they had free access to 1.8 ml of sucrose, followed by free access to food until *ex vivo* experiments. (iii) Paired, sucrose was delivered into a port 1 s after the onset of a tone, and the tone was terminated 400 ms after the animal entered the port to claim sucrose. The tone lasted for a maximum length of 30 s. If there was sucrose in the port during the onset of a tone (indicated by the absence of a port entry after the previous tone), then no sucrose was delivered in that trial. Mice could receive up to 120 sucrose deliveries and the conditioning session lasted about 4 h after which they had free access to food until *ex vivo* experiments. Behavioural performance from the second half of the conditioning session was used to assess performance and mice that met learning criterion (see Methods) were categorized in the learner group and the rest of the mice were categorized in the non-learner group. One day after conditioning, BLA neurons identified as either NAc or CeM projectors (retrobead positive) were recorded with whole-cell patch-clamp in *ex vivo* brain slices. *Ex vivo* data from both NAc and CeM projectors were collected from the following 7 groups: (1) naive ($n = 12$ mice, 9 for NAc pr., 7 for CeM pr.); (2) unpaired fear ($n = 13$ mice, 7 for NAc pr., 9 for CeM pr.); (3) paired fear ($n = 10$ mice, 7 for NAc pr., 7 for CeM pr.); (4) naive food restricted ($n = 11$ mice, 7 for NAc pr., n=8 for CeM pr.); (5) unpaired ($n = 11$ mice, 7 for NAc pr., 5 for CeM pr.); (6) reward paired non-learner ($n = 9$ mice, 5 for NAc pr., 6 for CeM pr.); and (7) reward paired learner ($n = 10$ mice, 8 for NAc pr., 7 for CeM pr.) groups. The n values indicated here are the number of mice used in Fig. 1. Data from groups 1–5 and 7 are shown in Fig. 1 and Extended Data Figs 1, 2 and 3. Data from group 6 is shown only in Extended Data Fig. 3. **f**, Time course of percentage freezing for the paired fear group. Percentage freezing was estimated during the shock-predictive tone (excluding the final 2 s, where the foot shock was delivered). **g**, Average normalized histogram of port entries relative to the onset of the tone predicting sucrose delivery for mice that learned the conditioned–unconditioned stimulus association (learners, $n = 11$) and mice that did not (non-learners, $n = 17$; see Extended Data Fig. 3). Mice in the paired reward conditioning group were deemed learners if the number of port entries in the post-conditioned-stimulus period (1 to 8 s relative to conditioned stimulus onset, black line) were determined as significantly higher than the number of port entries in the pre-conditioned stimulus period (−8 to −1 s relative to conditioned stimulus onset, grey line) using a one-sided Wilcoxon rank sum test ($P < 0.001$).
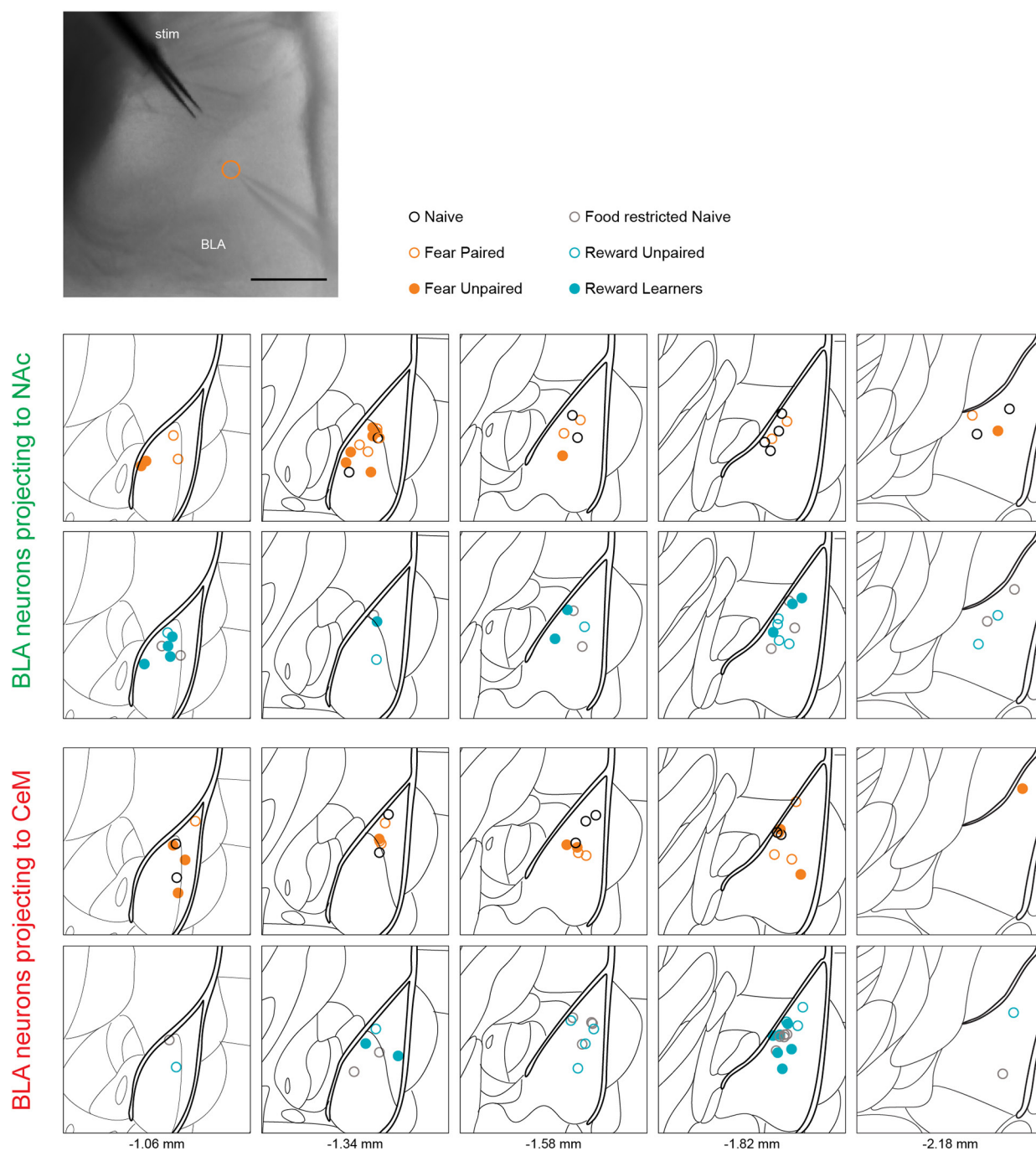
Extended Data Figure 2 | Location of BLA projectors recorded and analysed for each experimental group in Fig. 1. Top, representative DIC image showing the location of the stimulation electrode around a bundle of fibres of the internal capsule and a neuron recorded in the BLA (at the tip of the micropipette). The location of the recorded cell is indicated by an orange open circle. Scale bar, 200 μm. Bottom, atlas schematics (1.5 mm × 1.5 mm) showing BLA at various anteroposterior positions relative to Bregma. Each circle represents the location of a neuron from which the AMPAR/NMDAR ratio was acquired (Fig. 1). NAc projector locations are summarized in rows 1 and 2 and CeM projector locations are summarized in rows 3 and 4. Colour of the circle represents the conditioning group of the animal from which the AMPAR/NMDAR ratio was acquired.
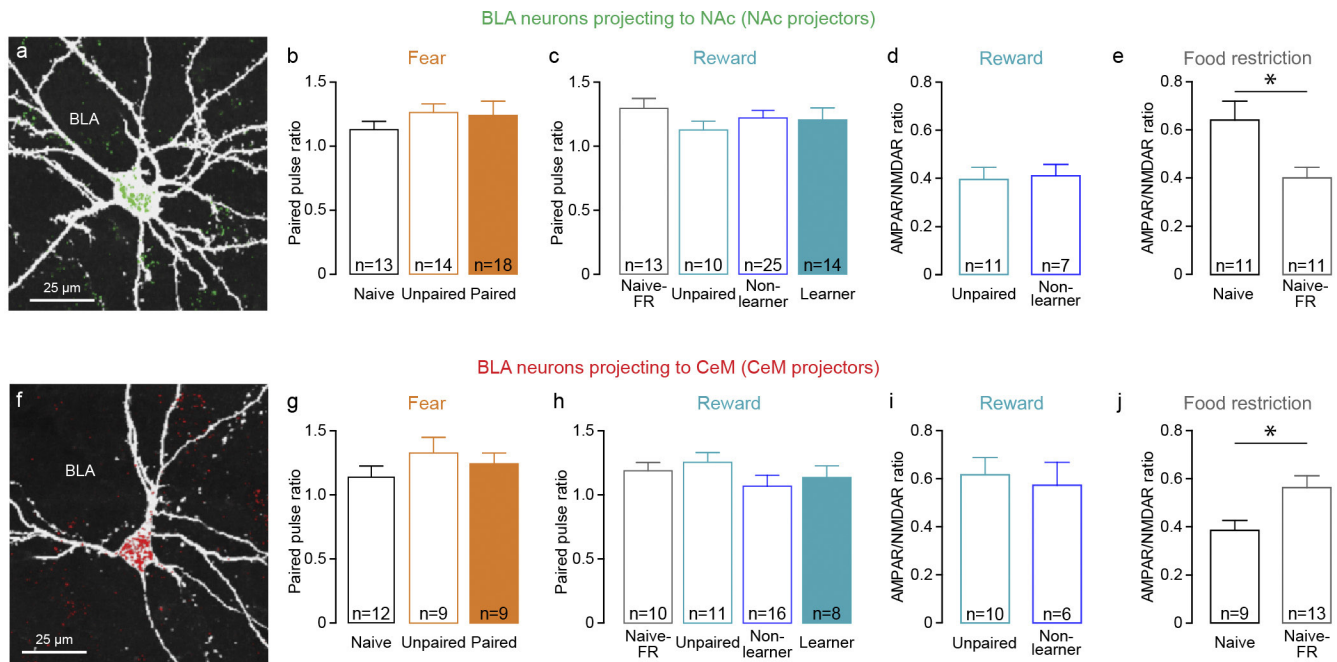
Extended Data Figure 3 | Paired-pulse ratio and AMPAR/NMDAR ratio in non-learners and food-restricted naive animals. a, Confocal image of a representative retrobead-positive neuron recorded in BLA after injection of retrobeads into NAc. This cell was recorded in an *ex vivo* slice, filled with biocytin and stained with streptavidin-CF405, pseudo-coloured white. **b,** In NAc projectors, the ratio of EPSC amplitude in response to paired-pulse stimulation (50 ms inter-pulse interval) of internal capsule inputs to the BLA was not related to experimental conditions of fear (one-way ANOVA, $F_{2,44} = 0.5209$, $P = 0.5978$). **c,** Paired-pulse ratio of EPSC amplitude was not related to experimental conditions of reward (one-way ANOVA, $F_{3,61} = 0.5868$, $P = 0.6261$). **d,** AMPAR/NMDAR ratio of internal capsule inputs onto NAc projectors in mice with unpaired tone and sucrose presentations (unpaired) and mice that did not learn the cue-reward association (non-learner) were not different (unpaired *t*-test $t_{16} = 0.180$, $P = 0.8595$). Both groups of mice received the same amount of total sucrose. **e,** AMPAR/NMDAR ratio on NAc projectors is significantly decreased by food restriction in naive mice (unpaired *t*-test, $t_{20} = 2.626$, $P = 0.0162$). **f,** Confocal image of a representative retrobead-positive neuron recorded in BLA after retrobead injection in CeM. **g,** Paired-pulse ratio of EPSC amplitude onto CeM projectors is not related to experimental conditions of fear (one-way ANOVA, $F_{2,29} = 0.9040$, $P = 0.4169$). **h,** Paired-pulse ratio of EPSC amplitude is not related to experimental conditions of reward (one-way ANOVA, $F_{3,44} = 0.9770$, $P = 0.4129$). **i,** AMPAR/NMDAR ratio on CeM projectors is similar in unpaired reward and paired reward non-learner mice (unpaired *t*-test $t_{14} = 0.381$, $P = 0.7090$). **j,** AMPAR/NMDAR ratio of internal capsule inputs onto CeM projectors is significantly increased by food restriction in naive mice (unpaired *t*-test $t_{20} = 2.526$, $P = 0.0201$). Results show mean and s.e.m.

a

NAc

ac

1.42 mm

NAc

ac

1.18 mm

ac

NAc

0.98 mm

b

LV

0.98 mm

c

CeM

BLA

- 0.82 mm

CeM

BLA

- 0.94 mm

CeM

BLA

- 1.06 mm

d

-0.94 mm

e

BLA

-1.22 mm

BLA

-1.34 mm

BLA

-1.46 mm

f

LV

-1.46 mm

BLA

-1.58 mm

BLA

-1.70 mm

BLA

-1.82 mm

g

-1.22 mm

BLA

-2.06 mm

○ RV-ChR2-Venus injection in NAc ───○─── Fiber placement for animals with RV-ChR2-Venus in NAc

◇ RV-ChR2-Venus injection in CeM ───◇─── Fiber placement for animals with RV-ChR2-Venus in CeM

□ RV-Venus injection in NAc/CeM ───□─── Fiber placement for animals with RV-Venus in NAc/CeM

● RV-ChR2-Venus injection of the animal in b ───●─── Fiber placement for animal in f

◆ RV-ChR2-Venus injection of the animal in d ───◆─── Fiber placement for animal in g

Extended Data Figure 4 | Histological verification of viral injection site and fibre placement for photostimulation experiments used in Fig. 2. a, Center of the rabies virus injection in NAc for the animals tested in intra-cranial self-stimulation (ICSS) and real-time place avoidance (RTPA) paradigms (Fig. 2a–e). Rabies virus (RV)-ChR2–Venus injections are denoted with green circles, and RV-Venus injections are indicated with grey squares. **b,** Representative confocal image of viral expression in a mouse 6 days after RV-ChR2–Venus injection in NAc. Right panel, enlarged view of the brightest fluorescence point (white circle), corresponding to the filled green circle in **a. c,** Center of RV-ChR2–Venus (red diamonds) and RV-Venus (grey squares) injections in CeM of animals analysed in Fig. 2. **d,** Example of viral expression 6 days after RV-ChR2–Venus injection in CeM. Right panel, enlarged view of the brightest fluorescence point (white circle), corresponding to the filled red diamond in **c. e,** Optical fibre tip placements over BLA of animals with RV-ChR2–Venus injected in NAc (green circles), CeM (red diamonds) or RV-Venus in NAc or CeM (grey squares). Horizontal lines represent the thickness of the implanted fibre (300 μm). **f,** Representative confocal image showing optical fibre tip from a RV-ChR2–Venus injection in NAc, corresponding to the filled green circle in **e**. Region in the white rectangle is magnified in the right panel and shows rabies-virus-expressing NAc projectors. **g,** Representative optic fibre placement for RV-ChR2-Venus injection in CeM, corresponding to the filled diamond in (**e**). Right panel: enlarged image of the BLA, containing rabies-virus-expressing CeM projectors. Atlas schematic in **a**, **c** and **e** represent 1.5 mm × 1.5 mm of the brain and the corresponding anteroposterior coordinates relative to Bregma are specified below. Scale bars in **b**, **d**, **f** and **g** are 500 μm.

Legend markers:
- ○ CAV-Cre injection in NAc of mice with AAV$_5$-EF1α-DIO-NpHR-eYFP in BLA
- □ CAV-Cre injection in NAc / CeM of mice with AAV$_5$-EF1α-DIO-eYFP in BLA
- ◆ CAV-Cre injection in CeM of mice with AAV$_5$-EF1α-DIO-NpHR-eYFP in BLA

- —○— Fiber placement for animals with CAV-Cre in NAc and AAV$_5$-EF1α-DIO-NpHR-eYFP in BLA
- —◆— Fiber placement for animals with CAV-Cre in CeM and AAV$_5$-EF1α-DIO-NpHR-eYFP in BLA
- —□— Fiber placement for animals with CAV-Cre in NAc / CeM and AAV$_5$-EF1α-DIO-eYFP in BLA

Extended Data Figure 5 | Histological verification of viral injection site and fibre placement for photoinhibition experiments used in Fig. 3. a, Centre of canine adenovirus (CAV)-Cre injection into bilateral NAc of mice with AAV$_5$-EF1α-DIO-NpHR-eYFP (green circles) or AAV$_5$-EF1α-DIO-eYFP (grey squares) injected bilaterally into the BLA. This approach allows for selective expression of NpHR–eYFP/eYFP, in NAc-projecting BLA neurons. **b**, Representative confocal image of the CAV-Cre injection site in NAc. **c**, Center of CAV-Cre injection into CeM from both hemispheres of mice with AAV$_5$-EF1α-DIO-NpHR-eYFP (red diamonds) or AAV$_5$-EF1α-DIO-eYFP (grey squares) injected bilaterally into BLA. In these animals, CeM-projecting BLA neurons express NpHR–eYFP or eYFP, respectively. **d**, Confocal image of a representative CeM injection and NpHR–eYFP-expressing cells bodies in the BLA. **e**, Optical fibre tip placements over BLA from both hemispheres in animals injected with AAV$_5$-EF1α-DIO-NpHR-eYFP in BLA and CAV-Cre in NAc (green circles) or CeM (red diamonds), or AAV$_5$-EF1α-DIO-eYFP in BLA and CAV-Cre in NAc/CeM (grey squares). Horizontal lines represent thickness of the implanted fibre (300 μm). **f**, Representative confocal images of optic fibre placements over BLA from both hemispheres of an animal injected with CAV-Cre in NAc and AAV$_5$-EF1α-DIO-NpHR-eYFP in BLA. Note NpHR–eYFP-expressing NAc projectors in the BLA. Each atlas diagram and confocal image in **a**–**f** represents an area of 1.5 mm × 1.5 mm; anteroposterior stereotaxic coordinates relative to Bregma are specified to the left of each image.
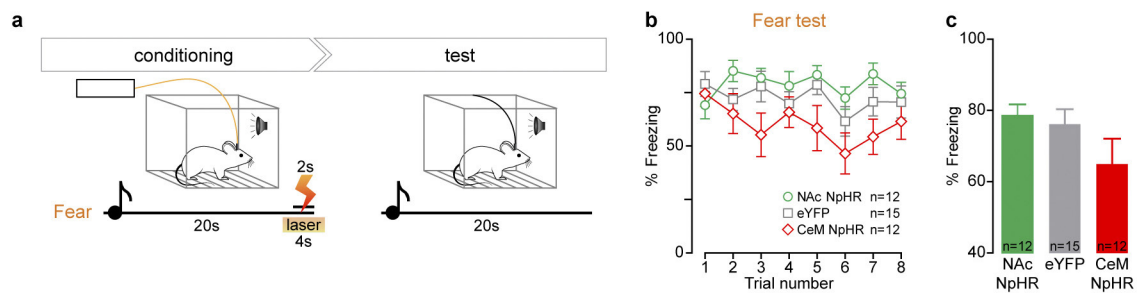
Extended Data Figure 6 | Tone-evoked freezing behaviour following inhibition of CeM or NAc projectors during auditory fear conditioning.
a, Experimental design. Mice were trained in an auditory fear conditioning paradigm, during which NAc or CeM projectors were selectively inhibited using a dual virus recombination approach (Fig. 3). On the day following conditioning, mice were exposed to eight presentations of the conditioned stimulus alone. They were tethered to a patch cable but no light was delivered.

b, Time course of percentage freezing in mice expressing NpHR in NAc projectors (green circles), CeM projectors (red diamonds), or expressing eYFP in NAc or CeM projectors (grey squares) was quantified for each trial. **c**, There was no significant difference in freezing behaviour in response to the conditioned stimulus among the three groups of mice on test day (one-way ANOVA, $F_{2,38} = 2.010$, $P = 0.1488$). Results show mean and s.e.m.
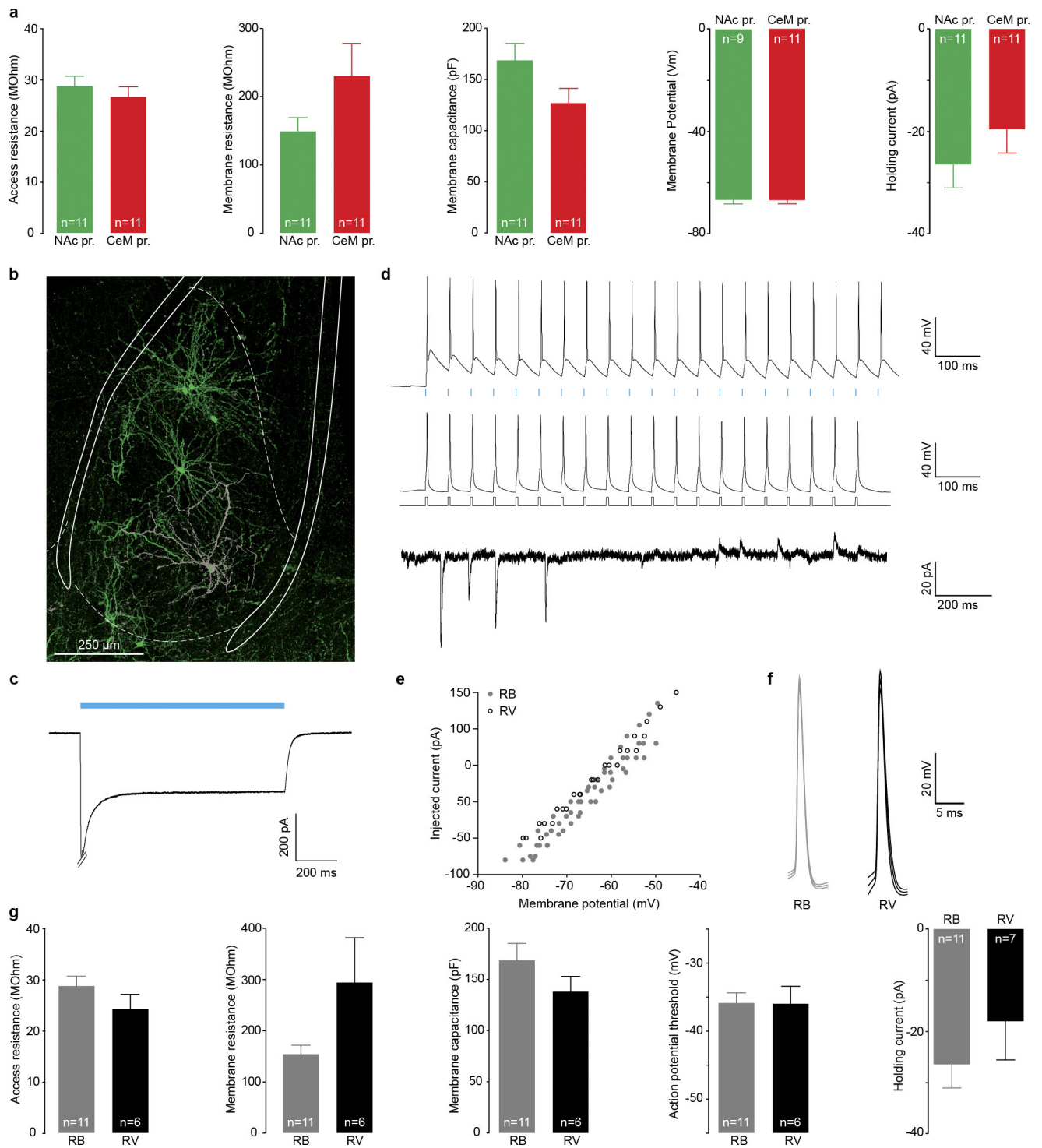
Extended Data Figure 7 | Membrane properties of retrobead-positive NAc/CeM-projecting BLA neurons and rabies-virus-expressing BLA neurons.
a, Access resistance, membrane resistance, and membrane capacitance were estimated from the current response of the cell to a 4 mV square voltage pulse using the Q-method[35]. Access and membrane resistance as well as the membrane capacitance and membrane potential were not significantly different between the two populations (unpaired t-tests: $t_{20} = 0.788$, $P = 0.4400$; $t_{20} = 1.599$, $P = 0.1256$; $t_{20} = 1.847$, $P = 0.0796$; and $t_{18} = 0.2521$, $P = 0.8038$, respectively). The holding current corresponds to the current injected to clamp the cell at -70 mV. This value was not significantly different between NAc and CeM projectors (unpaired t-test, $t_{20} = 1.046$, $P = 0.3079$). **b**, Confocal image of a BLA–NAc projectors expressing ChR2–eYFP transduced by rabies virus (RV) and recorded *ex vivo* in whole-cell patch-clamp. The cell was filled with biocytin during recording and stained with streptavidin-CF405 (in grey). **c**, Current response to a 1-s blue light pulse in a cell expressing rabies virus, 5 days after injection. **d**, Five days after viral injection, rabies-virus-expressing cells were able to respond with an action potential to every pulse of a 20 Hz light stimulation (5 ms pulses, top trace, blue line shows onset of light pulse). They also responded with an action potential to 250 pA, 5 ms current pulses injected at 20 Hz (middle trace). Rabies-virus-expressing cells also showed spontaneous post-synaptic excitatory and inhibitory currents (EPSCs and IPSCs, respectively) when clamped at -70 mV (bottom trace, 0 pA holding for this cell). **e**, Current/voltage curves are similar in retrobead (RB, grey circles, $n = 5$ cells) and rabies-virus-expressing cells (black circles, $n = 3$ cells). **f**, Average action potential for 11 retrobead-positive BLA–NAc projectors (grey) and six BLA–NAc projectors expressing rabies virus. **g**, Membrane properties of retrobead-positive versus rabies-virus-expressing neurons. None of the properties investigated were significantly altered in rabies-virus-expressing neurons (unpaired t-tests: access resistance, $t_{15} = 1.299$, $P = 0.2135$; membrane resistance, $t_{15} = 2.057$, $P = 0.0575$; membrane capacitance, $t_{15} = 1.215$, $P = 0.2430$; action potential threshold, $t_{15} = 0.0756$, $P = 0.9407$; holding current, $t_{16} = 1.002$, $P = 0.3314$). Results show mean and s.e.m.
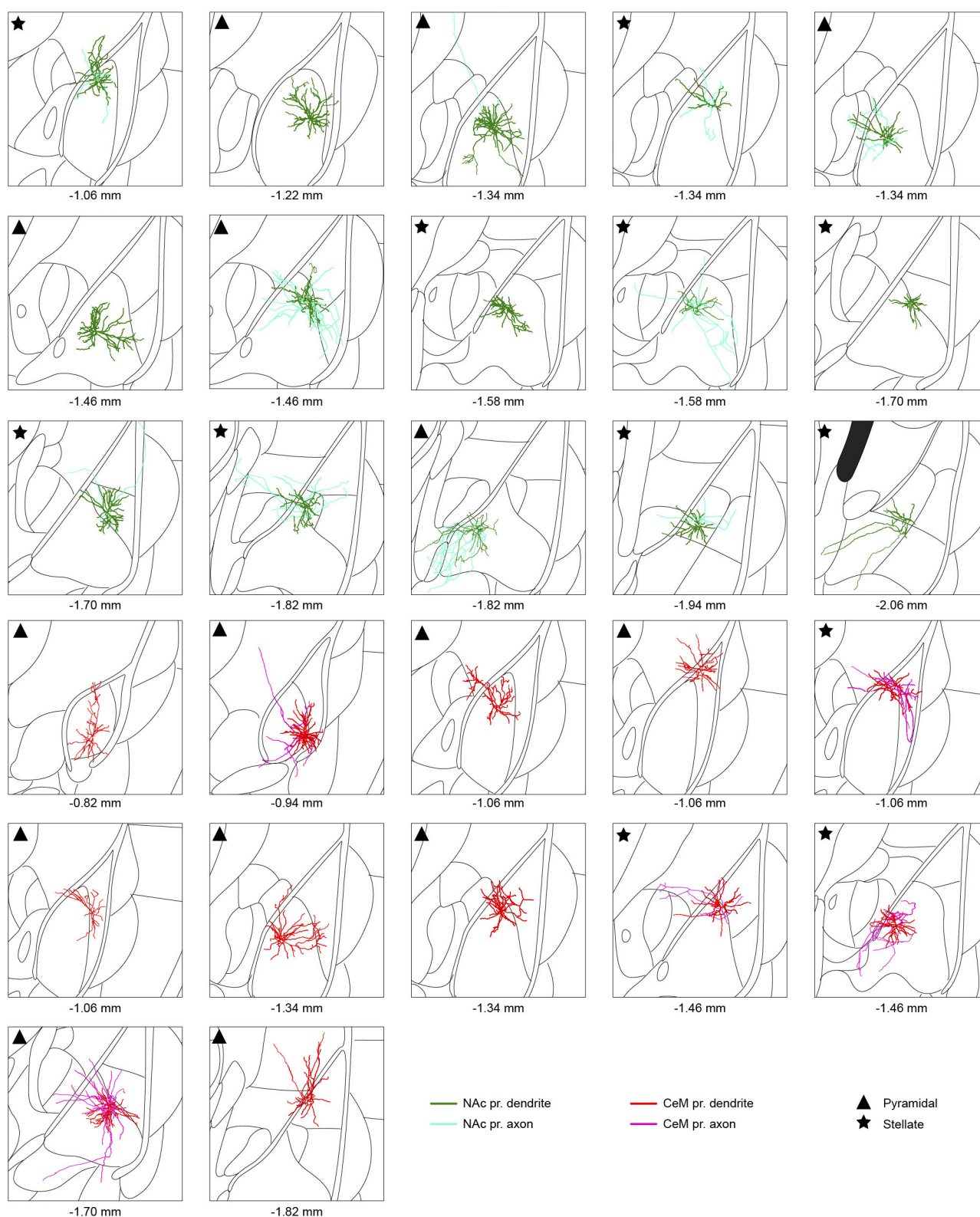
Extended Data Figure 8 | Morphological reconstructions of individual BLA neurons projecting to NAc or CeM. Morphological reconstructions of all neurons used for Sholl analysis performed by Imaris software (Fig. 4i). Classification of each neuron as pyramidal or stellate is indicated in the top left corner of each reconstructed neuron (triangle or star, respectively). Each atlas schematic represents 1.5 mm × 1.5 mm area and the corresponding anteroposterior stereotaxic coordinates (relative to Bregma) are shown below.
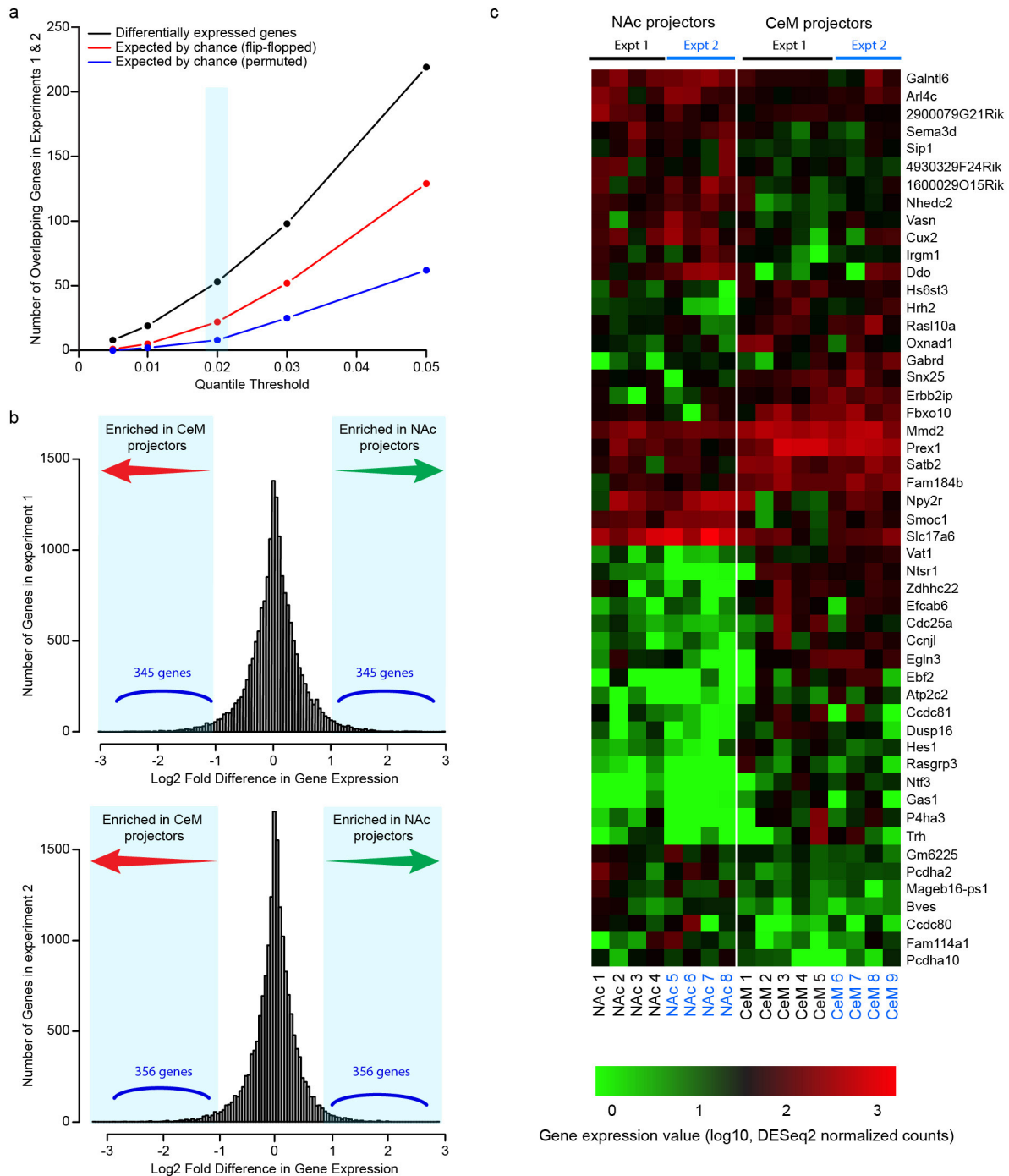
Extended Data Figure 9 | RNA-seq identification of candidate genes differentially expressed in NAc- and CeM-projecting BLA neurons.
a, Candidate differentially expressed genes were required to be enriched in only one group (either CeM or NAc projectors) in two independent experiments (NAc projectors collected from $n = 8$ mice; CeM projectors collected from $n = 9$ mice, total) at the indicated quantile fold-change threshold (light-blue column). One of the chance estimates ('flip-flopped', see Methods) is taken from genes that passed the quantile thresholds but were enriched in the opposite groups in the two experiments. Another chance estimate ('permuted', see Methods) is determined based on an analysis in which fold differences for each gene were permuted across genes within each of the two experiments before determining differential expression. A 0.02 quantile threshold was chosen to identify differentially expressed candidate genes in order to balance specificity and sensitivity, resulting in an estimated false discovery rate of 41.5%, calculated as the number expected by chance (flip-flopped) divided by the number of differentially expressed genes (see Extended Data Fig. 9c for candidate gene list). In Fig. 4k, a 0.01 quantile threshold was chosen to identify a more conservative list of differentially expressed candidate genes at a lower false discovery rate of 26.2%. **b**, Distribution of differentially expressed genes between NAc and CeM projectors from RNA-seq experiments 1 and 2 (see Methods). Light-blue shaded areas represent the 2nd and 98th percentiles of the distributions. **c**, RNA-seq heat map showing normalized expression levels of differentially expressed genes in NAc- and CeM-projecting BLA neurons. Differentially expressed genes were required to be enriched in either NAc or CeM projectors in two independent experiments (samples used in experiment 1 are indicated in black text below the heat map; experiment 2 samples are indicated in blue text) at a 0.02 quantile threshold (Extended Data Fig. 9a). Each RNA-seq library was prepared from 35–60 manually sorted retrobead-labelled cells taken from the BLA.

LETTER

NIK1–mediated translation suppression functions as a plant antiviral immunity mechanism

Cristiane Zorzatto[1,2]*, João Paulo B. Machado[1,2]*, Kênia V. G. Lopes[1,2], Kelly J. T. Nascimento[1,2], Welison A. Pereira[1,2], Otávio J. B. Brustolini[1,2], Pedro A. B. Reis[1,2], Iara P. Calil[1,2], Michihito Deguchi[1,2], Gilberto Sachetto-Martins[2,3], Bianca C. Gouveia[1,2], Virgílio A. P. Loriato[1,2], Marcos A. C. Silva[2], Fabyano F. Silva[4], Anésia A. Santos[2], Joanne Chory[2,5] & Elizabeth P. B. Fontes[1,2]

Plants and plant pathogens are subject to continuous co-evolutionary pressure for dominance, and the outcomes of these interactions can substantially impact agriculture and food security[1–3]. In virus–plant interactions, one of the major mechanisms for plant antiviral immunity relies on RNA silencing, which is often suppressed by co-evolving virus suppressors, thus enhancing viral pathogenicity in susceptible hosts[1]. In addition, plants use the nucleotide-binding and leucine-rich repeat (NB-LRR) domain-containing resistance proteins, which recognize viral effectors to activate effector-triggered immunity in a defence mechanism similar to that employed in non-viral infections[2,3]. Unlike most eukaryotic organisms, plants are not known to activate mechanisms of host global translation suppression to fight viruses[1,2]. Here we demonstrate in *Arabidopsis* that the constitutive activation of NIK1, a leucine-rich repeat receptor-like kinase (LRR-RLK) identified as a virulence target of the begomovirus nuclear shuttle protein (NSP)[4–6], leads to global translation suppression and translocation of the downstream component RPL10 to the nucleus, where it interacts with a newly identified MYB-like protein, L10-INTERACTING MYB DOMAIN-CONTAINING PROTEIN (LIMYB), to downregulate translational machinery genes fully. LIMYB overexpression represses ribosomal protein genes at the transcriptional level, resulting in protein synthesis inhibition, decreased viral messenger RNA association with polysome fractions and enhanced tolerance to begomovirus. By contrast, the loss of *LIMYB* function releases the repression of translation-related genes and increases susceptibility to virus infection. Therefore, LIMYB links immune receptor LRR-RLK activation to global translation suppression as an antiviral immunity strategy in plants.

NIK1 was first identified as a virulence target of the begomovirus NSP[5,6]. For begomoviruses, a group of single-stranded DNA viruses that infect major crops, the success of infection relies not only on viral suppressors of RNA silencing[4] but also on the viral inhibitor, NSP, of the immune receptor, NIK1 (ref. 5). The NIK1 protein belongs to the same LRR-RLK subfamily as the well-characterized PAMP recognition co-receptor BRI1-ASSOCIATED RECEPTOR KINASE 1 (BAK1)[7,8]. NIK1 is involved in plant antiviral immunity[5], whereas BAK1 is required for plant immunity against bacteria, fungi and oomycetes through its interactions with multiple PAMP-recognition LRR-RLKs[9]. We have previously demonstrated that the activation of NIK1 kinase is induced by the phosphorylation of Thr 474 within the activation (A)-loop[10,11] (Supplementary Discussion 1). Apart from the identification of RPL10 as a downstream effector in NIK1-mediated antiviral immunity[12,13], mechanistic knowledge of the signalling pathway is lacking, and the molecular nature of the defence response remains unclear. In this study, we replaced the normal NIK1 receptor with the NIK1 phosphomimetic gain-of-function mutant T474D[11] in transgenic *Arabidopsis* lines to understand the molecular basis of the NIK1-mediated defence mechanism

(Extended Data Fig. 1a–c). Transgenic lines possessing the gain-of-function mutant T474D in the *nik1* knockout background[10] were challenged with infectious clones of the *Arabidopsis*-infecting begomovirus cabbage leaf curl virus (CaLCuV)[10,11]. Then, we compared the virus-induced and T474D-induced transcriptomes at 10 days post-inoculation (dpi). A global cluster analysis of the expressed sequences among the mock-treated and infected wild-type (Col-0), NIK1 and T474D lines (Supplementary Table 1) revealed that the transcriptomes of the infected wild-type and mock-inoculated T474D lines were most closely related; these samples clustered together with a high bootstrap probability and a high approximately unbiased *P* value (Fig. 1a), which suggests that the NIK1-mediated response and the response to begomovirus infection share similar mechanisms. These transcriptomes differed greatly from the NIK1 mock-inoculated transcriptome, indicating that virus infection activates the NIK1-mediated response. Moreover, the gain-of-function T474D mutant might be activated in a constitutive manner that allows it to support a sustained NIK1-mediated response, in contrast with the expression of the intact NIK1 receptor in the *nik1* genetic background. The transcriptome from NIK1-complemented lines clustered with the Col-0 mock-inoculated transcriptome.

We also employed these transgenic lines to assess the T474D-induced global variation in gene expression. Gene enrichment analyses of immune system category genes indicated that ectopic expression of T474D did not activate typical viral defences, such as salicylic acid signalling or virus-induced gene silencing (Supplementary Table 2, Extended Data Fig. 2a, b and Supplementary Discussion 2). Among the differentially expressed genes, we observed an overrepresentation of translational-machinery-related genes, which largely predominated the downregulated gene list (Extended Data Fig. 3a, red spots; Supplementary Tables 2 and 3). Using enrichment analysis, these downregulated genes included ribosomal genes and other components of the protein synthesis machinery. Therefore, T474D ectopic expression downregulates components of the translational machinery, suggesting that the constitutive activation of NIK1 might influence translation. To confirm that protein synthesis was impaired by constitutive activation of NIK1 in the T474D lines, we labelled leaf proteins *in vivo* with [^{35}S]Met in the control and *nik1* plants, as well as in the *NIK1*- (ref. 11), G473V/T474A- (ref. 11) and T474D-expressing lines (Fig. 1b and Extended Data Fig. 3b). The one-proportion statistical test indicated a significant decrease (12.8% for T474D line 4 (T474D-4) and 13% for T474D-6; *P* < 0.05) in the amount of newly synthesized protein in T474D-expressing leaves compared with wild-type and *NIK1*-expressing leaves. We also demonstrated that dexamethasone-inducible T474D expression for 8 h led to a higher inhibition of *de novo* protein synthesis in the transgenic lines (Fig. 1c and Extended Data Fig. 3c–e). In the dexamethasone-inducible lines, the expression of T474D significantly reduced polysome (PS) and monosome (NPS) fractions (12% total reduction) to a similar extent as it

[1]Departamento de Bioquímica e Biologia Molecular, National Institute of Science and Technology in Plant–Pest Interactions, Bioagro, Universidade Federal de Viçosa, 36570.000 Viçosa, Minas Gerais, Brazil. [2]National Institute of Science and Technology in Plant-Pest Interactions, Bioagro, Universidade Federal de Viçosa, 36570.000 Viçosa, Minas Gerais, Brazil. [3]Departamento de Genética, Universidade Federal do Rio de Janeiro, 21944.970 Rio de Janeiro, Brazil. [4]Departamento de Zootecnia, Universidade Federal de Viçosa, 36570.000 Viçosa, Minas Gerais, Brazil. [5]Howard Hughes Medical Institute and Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.
*These authors contributed equally to this work.
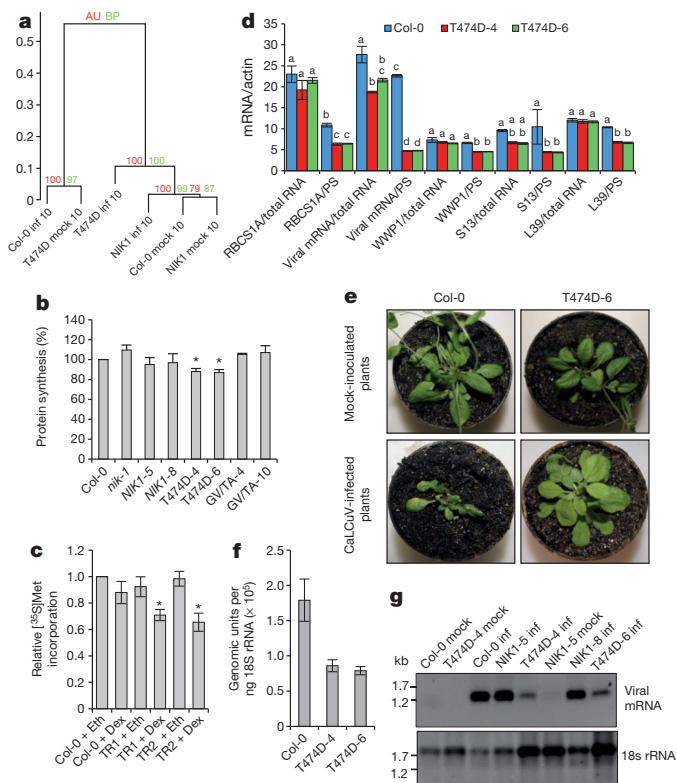
Figure 1 | Constitutive activation of the NIK1 receptor suppresses global host translation and confers tolerance to begomovirus. a, Ward hierarchical clustering of the gene expression data from the *Arabidopsis* infection experiments at 10 dpi. The dendrogram provides two types of *P* values: approximately unbiased (AU; red) and bootstrap probability (BP; green). inf, infected plants; mock, mock-inoculated plants. **b**, Global translation suppression by T474D expression. *In vivo* labelling of leaf proteins with [^{35}S]Met was performed in Col-0 and transgenic seedlings. The labelling percentage was also normalized to the leaf chlorophyll content (Extended Data Fig. 1d). Estimated proportion (in %) with respective error bars showing standard error of the mean (s.e.m.) from three independent experiments ($n = 3$). Asterisks indicate that the proportion is significantly ($P < 0.05$) different from 100% using the one-sided (less) chi-squared test. GV/TA, G473V/T474A double-mutant lines. **c**, Inhibition of global translation upon induction of T474D expression. An oestrogen-receptor-based chemical-inducible system was used to induce T474D expression in *Arabidopsis* seedlings, which were treated with 10 μM dexamethasone (Dex) or ethanol (Eth) for 8 h, and then were pulse labelled with L-[^{35}S]Met for 1 h. Tricarboxylic acid (TCA)-precipitable radioactivity in each sample was normalized to Col-0 + ethanol. Estimated proportion with respective error bars showing s.e.m. from three independent experiments ($n = 3$). Asterisks indicate that the proportion is significantly (*$P < 0.05$) different from 1 using the one-sided (less) chi-squared test. TR1 and TR2 are independently transformed lines harbouring T474D under the control of the dexamethasone-inducible promoter. **d**, Quantitation of total and PS-associated viral replication initiator protein (Rep) and host transcripts by qRT–PCR. The two-way analysis of variance (ANOVA) results were sliced by RNA fractions (total and PS). Different letters indicate significant differences among the RNA levels of the same gene by the TukeyHSD test ($P < 0.01$). mRNA/actin, mRNA of the gene normalized to actin mRNA. **e**, Symptoms associated with CaLCuV infection in wild-type and transgenic lines at 21 dpi. The figure shows representative samples from three independent experiments, each one with ten plants (biological replicates). **f**, Absolute quantitation of CaLCuV genomic units in infected lines at 14 dpi. An 18S ribosomal RNA (rRNA) target was run in parallel for normalizing the template load per reaction. Error bars, 95% confidence intervals based on bootstrap resampling replicates of four independent ($n = 4$) experiments. **g**, PS loading of viral mRNA in systemically infected leaves. PS-bound RNA was probed with viral Rep complementary DNA and 18S rDNA.

reduced PS and NPS RNA (13% reduction; Extended Data Fig. 4a–c). The loading of host mRNA (*RBCS*, *Arabidopsis thaliana* (*At*)*WWP1*, *S13* and *S39* genes) in actively translating PS fractions was significantly reduced in T474D-overexpressing lines compared to the wild-type line, although to a different extent (Fig. 1d, Extended Data Fig. 4d and Supplementary Discussion 3). Therefore, the activation of NIK1 reduces global levels of translation, but the effect may not be the same for all mRNAs. This downregulation of cytosolic translation might at least partially underlie the molecular mechanisms involved in NIK1-mediated antiviral defences.

To examine whether the constitutive activation of NIK1 was effective at controlling begomovirus infection, the transgenic lines were inoculated with CaLCuV DNA-A and DNA-B. The wild-type plants displayed typical symptoms of CaLCuV infection, whereas the symptoms in the T474D-expressing lines were greatly attenuated (Fig. 1e and Extended Data Fig. 5b). The symptomless CaLCuV infections of the T474D-expressing lines were associated with a delayed course of infection (Extended Data Fig. 5a) and a lower accumulation of viral DNA in the systemically infected leaves (Fig. 1f).

Because T474D expression caused downregulation of protein synthesis, we determined whether viral RNA translation was impaired in the T474D-expressing lines. We examined viral RNA transcripts in actively translating PS fractions prepared from infected leaves at 10 dpi (Extended Data Fig. 4e–g), when the total viral mRNA accumulation in the Col-0 and T474D-expressing lines was not very dissimilar (Fig. 1d and Extended Data Fig. 5c). We observed a significant reduction in the PS loading of viral mRNA in the T474D infected leaves compared with the wild-type infected leaves (Fig. 1d, g), indicating that the begomovirus was not capable of sustaining high levels of viral mRNA translation in the T474D-expressing lines.

Our data indicate that the translational suppression induced by NIK1 constitutive activation is associated with the downregulation of translational-machinery-related genes. Hence, NIK1-mediated
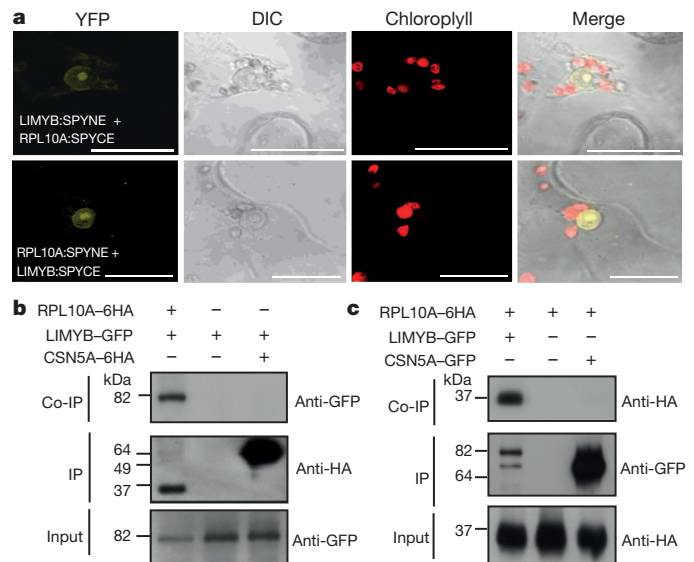
Figure 2 | LIMYB interacts with RPL10 in the nucleus. a, *In vivo* interaction between LIMYB and RPL10 by bimolecular fluorescence complementation (BiFC) analysis. The fluorescence (yellow fluorescent protein (YFP)) images were acquired using *Nicotiana tabacum* leaves co-expressing the 35S::RPL10-SPYNE + 35S::LIMYB-SPYCE and 35S::RPL10-SPYCE + 35S::LIMYB-SPYNE fusion proteins. They are representative samples from three independent biological repeats. Scale bars, 20 μm. DIC, differential interference contrast. **b**, LIMYB and RPL10 interaction *in planta*. The indicated constructs were expressed in *N. tabacum* leaves, and co-immunoprecipitation (Co-IP) assays were performed using an anti-haemagglutinin (HA) antibody. CSN5A is an unrelated protein used as a negative control. **c**, Co-immunoprecipitation assay was performed using an anti-GFP antibody.

nucleocytoplasmic trafficking of RPL10 may be linked to the regulation of gene expression. The previously identified extraribosomal functions of RPL10 associated with transcription factor regulation[14–16] may serve as potential targets for assessing this hypothesis. We used the two-hybrid system to search for RPL10 nuclear partners and isolated a MYB-domain-containing transcription factor, which was designated as LIMYB (Extended Data Fig. 6a–e). As a putative transcription factor, LIMYB localized in the nucleus of transiently or stably transformed plant cells (Extended Data Fig. 7a–i) and interacted with RPL10 in the nuclei of plant cells (Fig. 2a and Extended Data Fig. 6f, g). We further demonstrated that RPL10 and LIMYB interact *in vivo* using co-immunoprecipitation assays (Fig. 2b, c).

The function of LIMYB in NIK1-mediated antiviral signalling was examined using several different approaches. We first demonstrated that *LIMYB*, *RPL10* and *NIK1* are co-expressed in several organs (Extended Data Fig. 8). Then, we identified transfer DNA (T-DNA) insertion mutants (*limyb-32* and *limyb-82*) in the *LIMYB* gene (Extended Data Fig. 1e, f). We also prepared *LIMYB*-overexpressing lines (Extended Data Fig. 1g) and determined the LIMYB-induced global variation in gene expression compared to Col-0 leaves. Remarkably, the overexpression of *LIMYB*

resulted in a downregulation of translational-machinery-related genes similar to that induced by T474D expression (Extended Data Fig. 1h, red spots). We selected five ribosomal protein (RP) genes to confirm the deep-sequencing results for the *LIMYB*-overexpressing lines by quantitative polymerase chain reaction with reverse transcription (qRT–PCR) (Fig. 3a). *LIMYB* overexpression downregulated the expression of the selected RP genes but not of the unrelated gene (*AtWWP1*; Extended Data Fig. 9c). Conversely, in the *limyb-32* line, the RP genes were up-regulated (Fig. 3 and Extended Data Fig. 9a, d). Expression of *LIMYB* in *limyb* lines restored the wild-type expression of the RP genes (Extended Data Fig. 9b). As in the T474D-expressing lines, protein synthesis was slightly but significantly reduced in the *LIMYB*-overexpressing lines (Fig. 3c and Extended Data Fig. 4h). Because RPL10 functions in NIK1-mediated antiviral signalling and interacts with LIMYB, we examined whether RPL10 also controls the expression of RP genes. In *RPL10*-overexpressing lines (Extended Data Fig. 1i, j), the expression of RP genes but not of the unrelated *AtWWP1* gene was downregulated (Fig. 3d and Extended Data Fig. 9e). We then examined whether LIMYB, a putative transcription factor, binds to an RP promoter (RPL18) *in vivo*. We performed chromatin immunoprecipitation (ChIP) experiments
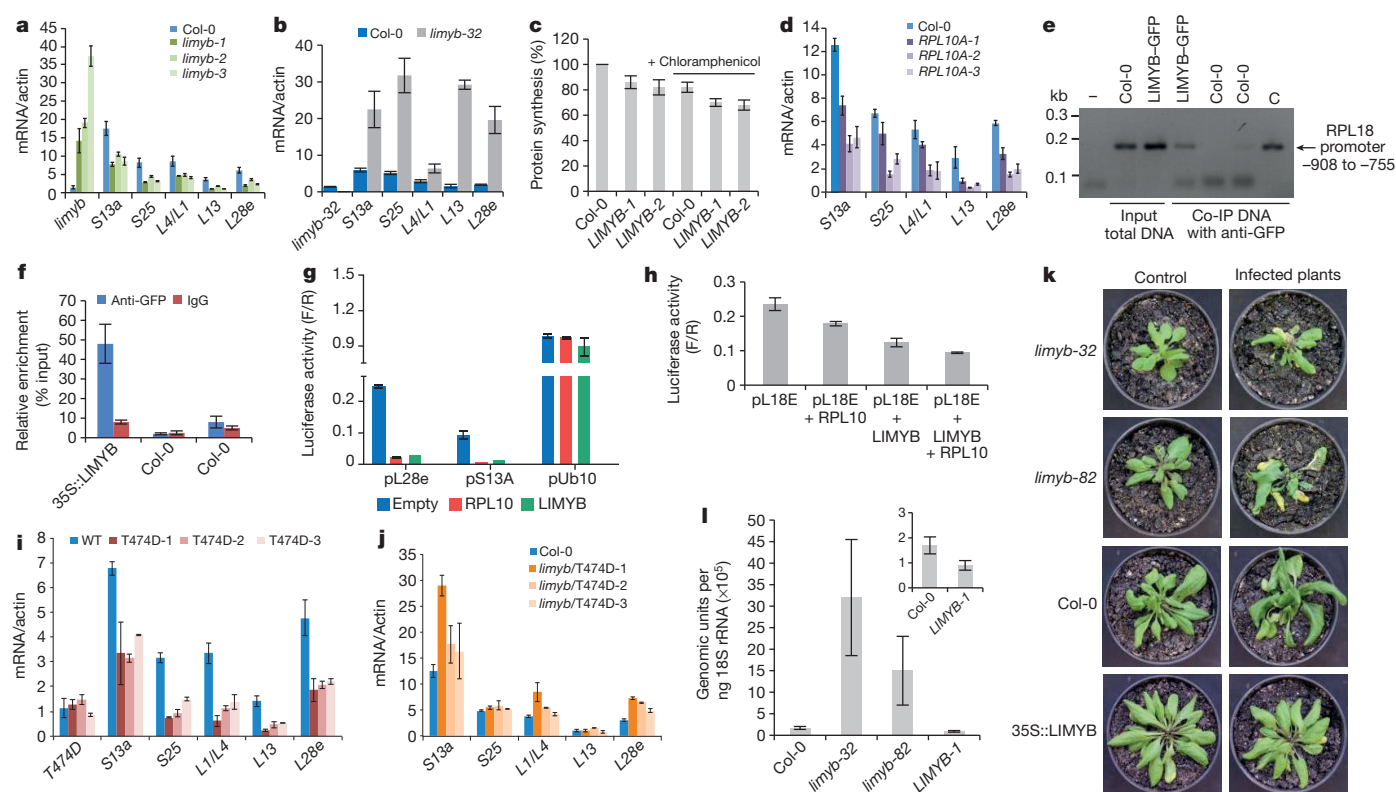
Figure 3 | LIMYB links NIK1 activation to the downregulation of RP genes and confers tolerance to begomovirus. a, *LIMYB* overexpression downregulates RP genes. RP gene expression was monitored by qRT–PCR of RNA from *LIMYB*-overexpressing leaves and Col-0. **b**, Induction of RP genes in the *limyb-32* mutant, monitored as in **a**. **c**, *LIMYB* overexpression suppresses global host translation. Col-0 and *LIMYB*-overexpressing seedlings were labelled with [35S]Met in the presence and absence of chloramphenicol, as described in Fig. 1b. **d**, RPL10 downregulates RP genes, monitored as in **a**. **e**, LIMYB binds to the RPL18e promoter *in vivo*. ChIP assay was performed with LIMYB–GFP-expressing leaves or wild type using anti-GFP antibodies. The co-immunoprecipitated (Co-IP) 150-bp fragment of the RPL18 promoter was detected by PCR. C, amplification of the control plasmid. **f**, ChIP-qPCR assay of 35S:LIMYB–GFP transgenic and wild-type (Col-0) seedlings. Samples were immunoprecipitated with anti-GFP antibody or IgG antibody. ChIP DNA was quantified with real-time PCR. **g**, LIMYB or RPL10 repress the RP promoter. *N. benthamiana* leaves were agro-infiltrated with plasmids carrying the prL28e-luciferase (pL28e), prS23A-luciferase (pS13A) or prUb10-luciferase (pUb10) genes in combination with either the *35S::LIMYB* or

35S:RPL10 DNA constructs. Luciferase activity was measured 48 h after agro-infiltration. F/R, firefly luciferase activity to *Renilla* activity ratio. **h**, LIMYB and RPL10 determine the full transcriptional repression of the RP promoter. *N. benthamiana* leaves were agro-infiltrated with the indicated combination of DNA constructs and luciferase activity was measured 48 h after infiltration. **i**, Downregulation of RP genes in T474D-expressing lines, monitored as in **a**. **j**, T474D requires the function of LIMYB to mediate RP gene suppression. The *limyb-32* (*limyb*) mutant was transformed with 35S::T474D, and the expression of RP genes was monitored as described in **a**. **k**, Symptoms associated with CaLCuV infection in Col-0 plants, the *LIMYB*-overexpressing line and *limyb* mutants at 21 dpi. Representative samples from three independent experiments, each one with ten plants (biological replicates). **l**, Absolute quantitation of CaLCuV genomic units in infected *LIMYB*-overexpressing lines and *limyb* mutants. Viral accumulation was determined by quantitative PCR at 14 dpi, as described in Fig. 1f. **a–d, f–j, l**, The respective 95% confidence interval limits were estimated based on bootstrap resampling replicates of three independent (*n* = 3) experiments and two technical repeats, except that in **g**, *n* = 2 with two technical repeats.

in which a 150 base pair (bp) RPL18 promoter was amplified from the precipitated DNA of *LIMYB*-expressing tissues but not of wild-type tissues (Fig. 3e). ChIP-qPCR showed that the 150-bp promoter fragment was significantly enriched in samples precipitated by anti-green fluorescent protein (GFP) antibody but not in samples pulled down from wild-type lines (Fig. 3f). These results suggest that LIMYB may function as a DNA-binding protein that associates with RP promoters *in vivo*.

To provide further evidence for the regulation of RP target genes by LIMYB and RPL10, we performed a luciferase transactivation assay in agro-infiltrated leaves. Consistent with the gene expression profile, LIMYB and RPL10 specifically repressed the L28e, S13A and L18E promoters but not an unrelated ubiquitin promoter (Fig. 3g, h). The co-transfection of *Arabidopsis* protoplasts with both LIMYB and RPL10 promoted increased repression of the L18E promoter compared to regulation of the gene reporter by individual expression of the transfactors. Collectively, these results indicate that LIMYB and RPL10 function as transcriptional repressors of common RP genes and that both transfactors are required for full regulation (Fig. 3h). Therefore, LIMYB and RPL10 may coordinately regulate common target promoters.

We also demonstrated that T474D expression downregulated the RP genes (Fig. 3i) but not *AtWWP1* (Extended Data Fig. 9f), confirming the RNA-sequencing data (Extended Data Fig. 3a). As an additional control, we showed that the double-mutant NIK1 G4743V/T474A[12], which does not complement the enhanced susceptibility phenotype of the *nik1*-null alleles (Extended Data Fig. 5a), also does not affect RP genes (Extended Data Fig. 9g). Nevertheless, the loss of *LIMYB* function prevented the T474D-mediated downregulation of the RP genes in both the *limyb-32* transgenic lines that stably expressed T474D (Fig. 3j and Extended Data Fig. 9h, i) and the *limyb-32* protoplasts that transiently expressed T474D (Extended Data Fig. 10). These results genetically link *LIMYB* to the translation suppression portion of the NIK1-mediated antiviral response.

We predicted that if the suppression of host translation was the basis for the begomovirus-tolerant phenotype of the T474D-expressing lines, *LIMYB* overexpression would also be effective against CaLCuV and the loss of *LIMYB* function would further debilitate the plant upon begomovirus infection. Col-0, *limyb-32*, *limyb-82* and *LIMYB*-overexpressing lines were inoculated with infectious clones of CaLCuV DNA-A and DNA-B (Fig. 3k). Both Col-0 and *limyb* lines developed typical CaLCuV symptoms, although to different extents. The disease symptoms varied in severity from extreme stunting and leaf distortion with severe chlorosis in the *limyb-32* and *limyb-82* leaves to mild leaf distortion and moderate chlorosis in Col-0. The course of infection in the *limyb* leaves was accelerated compared to that in the Col-0 plants (Extended Data Fig. 5d) and the *limyb-32* and *limyb-82* systemic leaves accumulated higher levels of viral DNA than Col-0 (Fig. 3l). Therefore, the loss of *LIMYB* function recapitulated the enhanced begomovirus susceptibility phenotype of the *nik1*-null alleles, as would be expected from a downstream component of the NIK1-mediated antiviral defence. By contrast, the *LIMYB*-overexpressing lines did not develop symptoms, displayed delayed infection and accumulated a lower level of viral DNA in their systemic leaves, resembling the tolerant phenotype of the T474D-overexpressing lines. We also found that *LIMYB* overexpression did not induce salicylic-acid-signalling marker genes or typical defence responses to viral infection (Extended Data Fig. 2c and Supplementary Table 2). Collectively, these results further support the notion that the inhibition of host translation observed in the T474D and *LIMYB* lines may be an effective mechanism exploited by plant cells to fight begomovirus infection. Therefore, the demonstration that immune-receptor-mediated defence signalling controls translation in plant cells may represent a new paradigm for antiviral defences in plants (Supplementary Discussion 4).

1. Pumplin, N. & Voinnet, O. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Rev. Microbiol.* **11**, 745–760 (2013).
2. Mandadi, K. K. & Scholthof, K. B. Plant immune responses against viruses: how does a virus cause disease? *Plant Cell* **25**, 1489–1505 (2013).
3. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
4. Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D. & Mansoor, S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Rev. Microbiol.* **11**, 777–788 (2013).
5. Santos, A. A., Lopes, K. V. G., Apfta, J. A. C. & Fontes, E. P. B. NSP-interacting kinase, NIK: a transducer of plant defence signalling. *J. Exp. Bot.* **61**, 3839–3845 (2010).
6. Mariano, A. C. *et al.* Identification of a novel receptor- like protein kinase that interacts with a geminivirus nuclear shuttle protein. *Virology* **318**, 24–31 (2004).
7. Sakamoto, T. *et al.* The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRII-RLK subfamily in antiviral defense. *BMC Plant Biol.* **12**, 229 (2012).
8. Shiu, S. H. & Bleecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
9. Chinchilla, D., Shan, L., He, P., de Vries, S. & Kemmerling, B. One for all: the receptor-associated kinase BAK1. *Trends Plant Sci.* **14**, 535–541 (2009).
10. Fontes, E. P. B., Santos, A. A., Luz, D. F., Waclawovsky, A. J. & Chory, J. The geminivirus NSP acts as virulence factor to suppress an innate transmembrane receptor kinase-mediated defense signaling. *Genes Dev.* **18**, 2545–2556 (2004).
11. Santos, A. A., Carvalho, C. M., Florentino, L. H., Ramos, J. J. O. & Fontes, E. P. B. Conserved threonine residues within the A-loop of the receptor NIK differentially regulate the kinase function required for antiviral signaling. *PLoS ONE* **4**, e5781 (2009).
12. Carvalho, C. M. *et al.* Regulated nuclear trafficking of rpL10A mediated by NIK1 represents a defense strategy of plant cells against virus. *PLoS Pathog.* **4**, e1000247 (2008).
13. Rocha, C. S., Santos, A. A., Machado, J. P. B. & Fontes, E. P. B. The ribosomal protein L10/QM-like protein is a component of the NIK-mediated antiviral signaling. *Virology* **380**, 165–169 (2008).
14. Oh, H. S., Kwon, H., Sun, S. K. & Yang, C.-H. QM, a putative tumor suppressor, regulates proto-oncogene c-Yes. *J. Biol. Chem.* **277**, 36489–36498 (2002).
15. Imafuku, I. *et al.* Presenilin 1 suppresses the function of c-Jun homodimers via interaction with QM/Jif-1. *J. Cell Biol.* **147**, 121–134 (1999).
16. Monteclaro, F. S. & Vogt, P. K. A Jun-binding protein related to a putative tumor suppressor. *Proc. Natl Acad. Sci. USA* **90**, 6726–6730 (1993).

Author Contributions C.Z. and J.P.B.M. co-wrote the manuscript and performed most of the experiments related to LIMYB isolation and characterization. K.V.G.L. performed the T474D-related experiments. K.J.T.N. performed qPCR for viral DNA. W.A.P. generated the RPL10 constructs and conducted plant transformation. O.J.B.B. and F.F.S. performed the bioinformatic analysis of the RNA-sequencing data and the statistical analysis of the data. M.D. performed the infectivity assays. B.C.G. conducted qRT–PCR. P.A.B.R. and I.P.C. performed the protein synthesis assays. V.A.P.L. performed complementation assays. G.S.-M. performed the bimolecular fluorescence complementation experiments. M.A.C.S. performed the tissue expression experiments. A.A.S. constructed the mutant proteins and designed the infectivity assays. J.C. conceived the experiments and edited the final draft. E.P.F.B. co-wrote the manuscript, designed the experiments and directed the project.

Author Information RNA-sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE56922. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.P.B.F. (bbfontes@ufv.br).

METHODS

Plasmid constructs. The clone pK7F-NIK1T474D was described previously[11]. This clone harbours a *GFP* gene fused in frame after the last codon of the respective mutant cDNA under the control of the CaMV 35S promoter. In the mutant cDNA T474D, the threonine residue at position 474 within the activation loop of NIK1 was replaced with an aspartate residue. The clones rpL10AST-pDONR201 (pUFV608), rpL10ANS-pDONR201 (pUFV609), rpL10AST-pDONR207 (pUFV900) and rpL10ANS-pDONR207 (pUFV901), which harbour the RPL10A coding region either with (ST) or without (NS) a translational stop codon inserted into the entry vectors pDONR201 or pDONR207, have also been previously described[12]. The RPL10A coding region was transferred from these entry clones to the yeast expression vectors pDEST32 and pDEST22, generating the clones pBD-RPL10A (pUFV1422) and pAD-RPL10 (pUFV785) as GAL4 binding domain (BD) or activation domain (AD) fusions. For BiFC, the RPL10A coding region was transferred to the vectors SPY-NE-GW and SPY-CE-GW. The resulting clones, pSPYNE-RPL10A (pUFV1652) and pSPYCE-RPL10A (pUFV1653), harboured the RPL10A coding region fused to either the N-terminal (NE) region of YFP or the YFP C terminus (CE), respectively, under the control of the 35S promoter. The clones pK7F-L10 and pYFP-L10, which harbour RPL10A fused in frame with GFP after its last codon or with YFP before its first codon under the control of the 35S promoter, respectively, have been previously described[10].

The clone pYFP-NLS-L10, which harbours a YFP–L10 fusion with a nuclear localization signal (NLS) under the control of the 35S promoter, was constructed by first inserting the amplified coding region of RPL10A into the SacI site of the pGR vector. The resulting clone, pGR-L10, which contains the RPL10A coding region fused to a glucocorticoid receptor (GR) domain with an NLS at the 5′ end, was then used as the template for the amplification of the NLS-containing L10 fusion with specific primers. The resulting product was inserted by recombination into pDONR207 and transferred to 35S-YFP-casseteA-Nos-pCAMBIA1300, generating pYFP-NLS-L10.

For the induction of T474D expression, the mutant cDNA T474D was transferred by recombination from the pDON201-T474D entry vector to the destination vector pBAV150 (ref. 17). The resulting clone pBAV150-NIK1 T474D–GFP harbours the mutant T474D open reading frame (ORF) under the control of a dexamethasone-inducible promoter and fused to a C-terminal GFP tag.

***Arabidopsis* growth conditions and transformation.** The Columbia (Col-0) ecotype of *A. thaliana* was used as the wild-type control, and the *nik1* knockout line[10] was used for plant transformation. The *nik1* lines[10] were transformed with pK7F-NIK1T474D or pBAV150- NIK1 T474D-GFP using the floral dip method. Two independently transformed lines expressing the T474D transgene (T474D-4 and T474D-6) were selected for the infection assays. The transgenic lines ectopically expressing NIK1–GFP (*NIK1-5* and *NIK1-8*) or an inactive kinase, the double mutant of NIK1, and the G473V/T474A–GFP dead kinase (G473V/T474A-10 and G473V/T474A-8) have been previously described[11,12] (Supplementary Table 1).

RT–PCR and real-time RT–PCR analyses. Total RNA was extracted from *Arabidopsis* leaves using TRIzol (Invitrogen). RT–PCR assays were performed using gene-specific primers, as previously described[18]. For gene expression analysis by qRT–PCR, gene-specific primers were designed using Primer Express 3.0 (Life Technologies). Real-time RT–PCR assays were performed on an ABI7500 instrument (Life Technologies) using the SYBR Green PCR Master Mix (Life Technologies). The amplification reactions were as follows: 2 min at 50 °C, 10 min at 95 °C and 40 cycles of 94 °C for 15 s and 60 °C for 1 min. The variation in gene expression was quantified using the comparative Ct method ($2^{-\Delta\Delta Ct}$), and absolute gene expression was quantified using the $2^{-\Delta Ct}$ method. The values were normalized to endogenous actin and ubiquitin as control genes[11].

RNA sequencing and data analysis. The transgenic and wild-type lines were infected at the seven-leaf stage with CaLCuV, as described later. After 10 dpi or after 21 dpi, total RNA from systemically infected leaves, as diagnosed by PCR, and mock-inoculated leaves from wild-type, 35S::NIK1-5 and 35S::T474D-4 lines was isolated using TRIzol (Invitrogen). For the RNA-sequencing experiments, we used two biological replicates of a pool of ten plants at 10 dpi, when we detected high levels of viral DNA in systemic leaves but symptoms had not yet developed, or at 21 dpi, when symptoms were visible. The RNA-sequencing data were obtained using an Illumina Hi-seq 2000. The paired-end 100-bp protocol was used with the following quality filter parameters: 5 bases trimmed at the 3′ and 5′ ends of the reads, a minimum average quality of 30 phred score. The data were stored in a comma-separated value spreadsheet file, and differential gene expression (DGE) analysis was performed using the R/Bioconductor package edgeR[19]. The raw data were normalized using the TMM normalization factor[20]. The dispersion was estimated by the tagwise edger parameter. Differential expression was determined using the cut-off *P* value of 0.1 adjusted by the false discovery rate (FDR). The read mapping process was executed using the Bowtie program[21] with the cDNA data set retrieved from The *Arabidopsis* Information Resource (TAIR) database (http://www.arabidopsis.org), tenth release.

Gene ontology classification was performed using the R/Bioconductor packages GSEABase and GOstats. Clustering analysis was performed using the R package pvclust (Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling) using Ward's method[22], and heatmaps were generated using gplots. The results were stored in a relational database created in PostgreSQL, and a web interface was created using PHP to allow the database to be accessed and navigated (http://arabidopsisnik.inctipp.ufv.br).

***In vivo* labelling of leaf proteins.** *Arabidopsis* seedlings (300 mg) were incubated with 1 ml of nutrient solution containing 30 µg ml^{-1} dexamathosone, 10 µg ml^{-1} cycloheximide, 10 µg ml^{-1} puromycin or 25 µg ml^{-1} chloramphenicol for 1 h, 4 h or 8 h. After the incubation period, 20 µCi of [^{35}S]Met (EasyTag Protein Labelling Mix, [^{35}S]-, 2 mCi (74 MBq), Perkin Elmer) was added for 1 h at room temperature. To quantitate the incorporation of [^{35}S]Met into protein, aliquots of protein extracts were placed in 10% (w/v) TCA and incubated on ice for 30 min. The samples were filtered onto glass microfibre filters, and the filters were washed three times with 5 ml cold 5% (w/v) TCA and two times with 5 ml 95% ethanol. After drying, the filters were counted with a scintillation counter.

PS fractionation. PSs were fractionated from 500 mg of 15-day-old *Arabidopsis* seedlings over sucrose gradients as described[23]. Fractions were collected manually from the top, and total RNA was extracted with phenol/chloroform/isoamyl alcohol, precipitated with isopropanol, and treated with DNase I. The specific transcripts were examined by northern blot analysis. For the infection assays, PSs from wild type, *NIK1-4*-overexpressing and T474D-6-overexpressing lines were isolated 10 dpi with infectious CaLCuV clones.

PS gradients of dexamethasone-inducible T474D seedlings were prepared from 8-day-old seedlings, treated or not with 30 µg ml^{-1} dexamathosone. After 8 h of dexamethasone treatment, the seedlings were dried, frozen and ground in liquid nitrogen. Cytoplasmic extracts were prepared from 350 mg of powder that were resuspended in the ice-cold extraction buffer. Cell debris were removed by centrifugation and 0.6 ml of the supernadant were loaded onto a 4.5 ml 15–55% sucrose density gradient and separated by ultracentrifugation. Fractions (16) of 310 µl were collected from the top. Fractions 2–7 correspond to density regions of complexes ≤80S (NPSs) and fractions 8–15 correspond to density regions containing small and large polysomes (PSs). Quantitation of NPSs and PSs from the absorbance profile was performed as described previously[24]. The quantitation of total RNA from the combined NPS and PS fractions was performed as described previously[25]. To quantify mRNA loading into PSs, total RNA isolated from combined NPS (2–7) and PS (9–15) fractions was extracted using Trizol (Invitrogen), precipitated with isopropanol and quantified by qRT–PCR, as described earlier. qRT–PCR on each fraction from the PS gradient of dexamethasone-inducible T474D seedlings was also performed for quantitation of the distribution of *AtWWP1* and *S39* transcripts after induction of T474D expression. Each sample was amplified with 1 cycle at 50 °C for 2 min, 95 °C for 10 min and 40 cycles at 95 °C for 15 s and 60 °C for 1 min. Values were normalized to actin.

CaLCuV inoculation and viral DNA accumulation. *A. thaliana* plants at the seven-leaf stage were inoculated with plasmids containing partial tandem repeats of CaLCuV DNA-A and DNA-B by biolistic delivery[26], and the course of infection was monitored as described previously[10]. We used an attenuated form of the virus in which the coat protein ORF in the CaLCuV DNA-A was interrupted by the introduction of a stop codon at amino acid position 47. The inoculated plants were transferred to a growth chamber and examined for symptom development (leaf necrosis, chlorosis, leaf epinasty, leaf curly, young leaf death and stunted growth). Total nucleic acids were extracted from systemically infected leaves, and viral DNA was detected by PCR with DNA-B CaLCuV-specific primers (566CLCVBFBR1v, 5′-GGCGTG GGGTATCTTACTC-3′, and 1253CLCVBRBR1c, 5′-GACATAGCATCGGACA TCC-3′) as well as the actin-specific primers as an endogenous control. In each experiment, 20 plants of each line (Col-0, *nik1* and *nik1*-expressing NIK1 mutant proteins) were inoculated with 2 µg of tandemly repeated DNA-A plus DNA-B per plant. The course of infection was examined using data from three independent experiments.

Viral DNA accumulation was measured by qPCR. The reactions were prepared in a final volume of 10 µl using the Fast SYBR Green Master Mix (Life Technologies) according to the manufacturer's instructions and analysed on a 7500 Real Time PCR System. Virus-specific primers were designed using Primer Express 3.0 (Life Technologies) and tested by conventional PCR using plasmids containing the complete DNA-B of each virus (10^6 copies per reaction). The genomic copies of CaLCuV were normalized against an internal control (18S rRNA) to consider template input variation between tubes. CaLCuV DNA was amplified with primers B-Fwd (5′-GGGCCTGGGCCTGTTAGT-3′) and B-Rvs (5′-ACGGAAGATGGG AGAGGAAGA-3′). In this case, the genomic unit refers to one copy of the DNA-B of CaLCuV. PCR reactions were run in parallel with primers 18S-Fwd (5′-TAA TTTGCGCGCCTGCTGCC-3′) and 18S-Rvs (5′-TGTGCTGGCGACGCATCA TT-3′) for the reference plant gene *18SRNA*. Standard curves were obtained by

regression analysis of the Ct values of each of the three replicates of a given dilution as a function of the log of the amount of DNA in each dilution. Standard curves containing viral DNA and host DNA served as references. Each sample was analysed in triplicate from at least two biological replicates.

Two-hybrid screening. The yeast reporter strain MaV203 (MATαleu2-3,112 trp1-901 his3200 ade2-101 gal4 gal80 SPAL10::URA3 GAL1::lacZ HIS3UAS GAL1::HIS3-LYS2 can1R cyh2R; Trp- Leu- Ura-) was transformed sequentially with pBD-L10 and 25 µg of the pEXAD502 cDNA library previously prepared[27] and approximately 5×10^6 transformants were screened for histidine protoprophy an β-galactosidase activity, as described[27,28].

LIMYB-based plasmid constructs. The At5g05800 cDNA (LIMYB), which was isolated based on its ability to bind to rpL10A in yeast, was amplified by PCR from the cDNA library vector using AT5G05800-specific primers, re-amplified with the primers AttB1-Fwd and AttB1-Rvs, and cloned by recombination into the entry vectors pDONR201 and pDONR207. The resulting products were then transferred to different destination vectors for expression in yeast (pDEST32 and pDEST 22) and plants (pK7FWG2, 35S-YFP-casseteA-Nos-pCAMBIA1300, pSPYNEGW and pSPYCEGW). The resulting clones, pBD-At5g05800ST (pUFV1903) and pAD-At5g05800ST (pUFV1480), enabled the expression of AT5G05800 (LIMYB) in yeast as GAL4 binding domain (BD) or activation domain (AD) fusions, respectively. The clone pAt5g05800-GFP (pUFV1395) harboured the AT5G05800 (LIMYB) coding region fused to the GFP N terminus under the control of the 35S promoter, whereas in pYFP-At5g05800 (pUFV1886), the At5g05800 coding region was fused to the YFP C terminus. In the resulting clones, pSPYNE-At5g05800 (pUFV1658) and pSPYCE-At5g05800 (pUFV1657), the At5g05800 cDNA was linked to the YFP NE and the YFP CE, respectively, under the control of the 35S promoter.

The HA epitope was fused to the At5g05800 C terminus using the triple Gateway system, which consisted of the vector p5′, in which the 35S (2×) promoter had been previously cloned, the clone pUFV1378 (At5g05800 cDNA without a translational stop codon in pDONR201) and the vector p3′, in which a HA (6×) epitope had been previously inserted. The At5g05800 cDNA was transferred by recombination into the destination vector pK7M34GW along with 2×35S promoter and the HA tail, yielding the clone p2×35S-At5g05800-6HA (pUFV1984), which enabled the expression of At5g05800 fused in frame to HA under the control of the 35S promoter in plants. The same triple Gateway system was used to place the LIMYB fused in frame with Cherry or GFP under the control of the LIMYB promoter.

Transient expression in *N. benthamiana* leaves. To determine the subcellular localization of the proteins, *N. benthamiana* leaves were agro-infiltrated with pAt5g05800-GFP, pYFP-At5g05800 and pYFP-NAC81 (ref. 29), while for the co-immunoprecipitation assays, they were agro-infiltrated with pL10-GFP and p2×35S-At5g05800-6HA. *Agrobacterium tumefaciens* strain GV3101 was used for agroinfiltration. *Agrobacterium* infiltration was performed in the leaves of 3-week-old *N. benthamiana*, as previously described[12].

Confocal microscopy. Approximately 72 h after agro-infiltration, 1-cm² leaf explants were excised, and the GFP and YFP fluorescence patterns were examined in epidermal cells with a ×40 or ×60 oil immersion objective and a Zeiss inverted LSM510 META laser scanning microscope equipped with an argon laser and a helium laser as excitation sources, as described[27].

Co-immunoprecipitation assays. The co-immunoprecipitation assay was performed using the µMACSTM Epitope Tag Protein Isolation Kit (MACS/Miltenyi Biotec) according to the manufacturer's instructions. Total protein extracts were prepared from *N. benthamiana* leaves that had been agro-infiltrated with pL10-GFP and/or p2×35S-At5g05800-6HA. At 48 h after infiltration, 200 mg of leaves were homogenized with 1 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 1% (v/v) Nonidet P-40) and incubated for 2 h with anti-GFP magnetic beads (MACS/Miltenyi Biotec) at 4 °C under gentle agitation. After extensive washing of the magnetic beads, the bound proteins were eluted using 50 ml of elution buffer pre-warmed to 95 °C. The immunoprecipitated proteins were separated by 10% (w/v) SDS–PAGE and immunoblotted with anti-HA (Miltenyi Biotec, catalogue number 130-091-972) or anti-GFP (Miltenyi Biotec, catalogue number 130-091-833) monoclonal antibodies. The reacting antibodies were detected using Signal West Pico Chemiluminescent Substrate (Thermo Scientific) according to the manufacturer's instructions.

BiFC. *N. benthamiana* leaves were agro-infiltrated as previously described for the transient expression experiments using the following combinations of recombinant plasmids: pSPYNE-At5g05800 + pSPYCE-rpL10A; pSPYCE-At5g05800 + pSPYNE-rpL10A; pSPYNE-At5g05800 + pSPYCE empty vector; pSPYNE-rpL10A + pSPYCE empty vector; pSPYCE-At5g05800 + pSPYNE empty vector; pSPYCE-rpL10A + pSPYNE empty vector. After incubation for 72 h, the leaf sectors were examined by confocal microscopy. YFP was excited at 514 nm using an argon laser, and YFP emission was detected using a 560–615 nm filter. The stability of the CE and NE YFP regions was monitored by immunoblotting of transfected leaf protein extracts with a polyclonal anti-YFP serum.

LIMYB- or RPL10A-expressing lines and T474D-expressing *limyb* lines. *Arabidopsis* Col-0 was transformed with pAt5g05800-GFP, generating the transgenic line LIMYB-1; with pYFP-At5g05800, generating the transgenic lines LIMYB-2 and LiMYB-3; and with pYFP-NLS-L10, generating the transgenic lines RPL10-1, RPL10-2 and RPL10-3. Homozygous seeds of the T-DNA insertion *limyb* mutants (Salk_032054C and Salk_082995C) were obtained from the *Arabidopsis* Biological Resource Center. The *limyb-32* mutant was transformed with the gain-of-function mutant T474D, generating the transgenic lines *limyb*/T474-1, *limyb*/T474D-2 and *limyb*/T474D-3. Primary transformants were selected using the appropriate antibiotic (50 µg ml^{-1} kanamycin or 30 µg ml^{-1} hygromycin), and the stable incorporation of the transgene in the plant genome was evaluated by PCR with gene-specific primers. The expression of the transgenes was monitored by qRT–PCR.

LIMYB promoter construct and GUS histochemical assay. Approximately 1 kb of the 5′ flanking sequences of NIK1, RPL10A and At5g05800 (LIMYB) were amplified from *Arabidopsis* genomic DNA using Taq Platinum and specific oligonucleotides and inserted into the cloning vector pCR8/GW/TOPO (Invitrogen). The promoter sequences were then transferred by recombination into the destination vector pMDC162. The resulting clones, pNIK1-MDC162, pL10A-MDC162 and pAt5g05800-MDC162 (pUFV 1892), harboured the respective promoter sequences of the three genes fused to the β-glucuronidase (GUS) reporter gene and were used to transform *Arabidopsis* Col-0 plants. GUS expression was assayed histochemically.

Luciferase reporter gene assay. The At1g29970 (RPL18A) promoter was isolated from *Arabidopsis* DNA by PCR and cloned into pDONR-P4-P1R, generating pUFV2155. The clone LUCF-term-pDON221 (pUFV 2132) harbours the firefly luciferase cDNA in the entry vector pDON221, whereas the clone 2×35S-RLUCF-pDONR-P2R-P3 (pUFV 2131) contains *Renilla* luciferase cDNA under the control of a 2× 35S promoter. The RPL18A promoter was transferred from pUFV2155 to the destination vector pK7m34GW (pUFV 1918), along with 2×35S-RLUCF-pDONR-P2R-P3 (pUFV 2131) and LUCF-term-pDON221 (pUFV 2132), by triple recombination. The resulting clone, prAt1g29970-Lucif-term-2X35S RLucif pH7M34GW (pUVF2231), harboured the firefly luciferase cDNA under the control of the rpL18A promoter, as well as the *Renilla* luciferase cDNA under the control of a 2× 35S promoter. The same procedure was used to clone the firefly luciferase cDNA under the control of the RPL28e, RPS13A and ubiquitin promoters for luciferase transactivation assays.

N. benthamiana leaves were agro-infiltrated with *A. tumefaciens* GV3101 strains carrying the following combinations of DNA constructs: At1g29970-Lucif-term-2×35S RLucif pH7M34GW; At1g29970-Lucif-term-2X35S RLucif pH7M34GW + AT5G05800NS-pK7FWG2; At1g29970-Lucif-term-2X35S RLucif pH7M34GW + AT5G05800NS-pK7FWG2 + rpL10ANS-pK7FWG2; and At1g29970-Lucif-term-2×35S RLucif pH7M34GW + rpL10ANS-pK7FWG2. Forty-eight hours after infiltration, 200 mg of leaf tissue was harvested for total protein extraction. Luciferase activity was assayed with the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

Protoplast preparation from *A. thaliana* leaves and transient expression assays. Protoplasts were prepared from 15-day-old *Arabidopsis* seedlings as previously described[30], except that digestion was initiated for 30 s under vacuum and then prolonged for 6 h with agitation at 80 r.p.m. The transient expression assays were performed by electroporation (250 V, 250 µF) 10 µg of the expression cassette DNA and 30 µg of the sheared salmon sperm DNA into 2×10^5–5×10^6 protoplasts in a final volume of 0.8 ml. The protoplasts were diluted into 8 ml of MS medium supplemented with 0.2 mg ml^{-1} 2,4-dichlorophenoxyacetic acid and 0.8 M mannitol at pH 5.7. After 36 h of incubation in the dark, the protoplasts were washed with 0.8 M mannitol plus 20 mM MES at pH 5.7 and frozen in liquid N_2 until further use.

Statistical analyses. All statistical analyses (including gene expression clustering, qRT–PCR and protein synthesis data) were performed in R.

Cluster analysis. Cluster analysis of the RNA-sequencing data was used to classify the treatments (mock-inoculated and infected wild type, T474D and NIK1) according to similarities in the profiles of genome-wide expression at 10 and 21 dpi. The uncertainty of clustering results caused by sampling variations was verified by the probability-based cluster analysis, which was implemented using the pvclust package of R software. In this context, the bootstrap (BP) and approximately unbiased (AU) probability are used to validate the reported cluster. The BP and AU values are the percentage that a given cluster appears in the bootstrap and multiscale bootstrap replicates, respectively.

RNA-sequencing differential gene expression analysis. The edgeR package of R/bioconductor software was used to carry out the gene expression analysis. This package assumes negative binomial distribution for the read counts, and the used normalization is given by the TMM (trimmed mean of *M* value) method. The significance of differential gene expression was reported in terms of *q* values (FDR-adjusted *P* values).

Enrichment analysis. Gene set enrichment analysis (GSEA) was used to uncover biological processes associated with sets of differentially expressed genes instead of

focusing on individual genes. This analysis was implemented in the GOstats package of R/bioconductor software by using the function GSEAGOHyperGParams, which uses the relationships among the gene ontology terms as extra information in the statistical inference of groups.

Percentage of protein synthesis analysis. Analysis of protein synthesis was based on the one-proportion one-sided (less) test using chi-squared distribution. The tested hypothesis was the 100% synthesized labelled protein for each treatment; thus, if a given treatment is significant, it indicates that the proportion (or percentage) is different from 100%.

Confidence interval analysis. Nonparametric bootstrap confidence intervals were used on our graphics to increase the accuracy of the confidence limits by an overlapping analysis[31]. This method was introduced as a nonparametric device for estimating standard errors and biases. It is an automatic algorithm for producing highly accurate confidence limits from a bootstrap distribution implemented with the 'boot' package in R software.

Phylogenetic analysis. MYB family sequences were retrieved from the Agris database (http://arabidopsis.med.ohio-state.edu). The alignment was performed by Maft aligner software[32], and the tree was built by Fasttree[33] software.

17. Vinatzer, B. A. *et al.* The type III effector repertoire of *Pseudomona syringae pv. syringae* B728a and its role in survival and disease on host and non-host plants. *Mol. Microbiol.* **62,** 26–44 (2006).
18. Delú-Filho, N. *et al.* A sucrose binding protein homologue from soybean affects sucrose uptake in transgenic tobacco suspension-cultured cells. *Plant Physiol. Biochem.* **38,** 353–361 (2000).
19. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
20. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).
21. Langmead, B. *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
22. Ward, J. H. Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58,** 236–244 (1963).
23. Kim, T. H., Kim, B. H., Yahalom, A., Chamovitz, D. A. & von Arnim, A. G. Translational regulation via 5′ mRNA leader sequences revealed by mutational analysis of the *Arabidopsis* translation initiation factor subunit eIF3h. *Plant Cell* **16,** 3341–3356 (2004).
24. Kawaguchi, R., William, A. J., Bray, E. A. & Bailey-Serres, J. Water deficit-induced translational control in *Nicotiana tabacum*. *Plant Cell Environ.* **26,** 221–229 (2003).
25. Kawaguchi, R., Girke, T., Bray, E. A. & Bailey-Serres, J. Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J.* **38,** 823–839 (2004).
26. Santos, A. A., Florentino, L. H., Pires, A. B. L. & Fontes, E. P. B. Geminivirus: biolistic inoculation and molecular diagnosis. *Methods Mol. Biol.* **451,** 563–579 (2008).
27. Florentino, L. H. *et al.* A PERK-like receptor kinase interacts with the geminivirus nuclear shuttle protein and potentiates viral infection. *J. Virol.* **80,** 6648–6656 (2006).
28. Carvalho, C. M. *et al.* A novel nucleocytoplasmic traffic GTPase identified as a functional target of the bipartite geminivirus nuclear shuttle protein. *Plant J.* **55,** 869–880 (2008).
29. Pinheiro, G. L. *et al.* Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene* **444,** 10–23 (2009).
30. Costa, M. D. L. *et al.* A new branch of endoplasmic reticulum-stress signaling and the osmotic signal converge on plant specific asparagine-rich proteins to promote cell death. *J. Biol. Chem.* **283,** 20209–20219 (2008).
31. DiCiccio, T. J. & Efron, B. Bootstrap confidence intervals. *Stat. Sci.* **11,** 189–228 (1996).
32. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
33. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490 (2010).
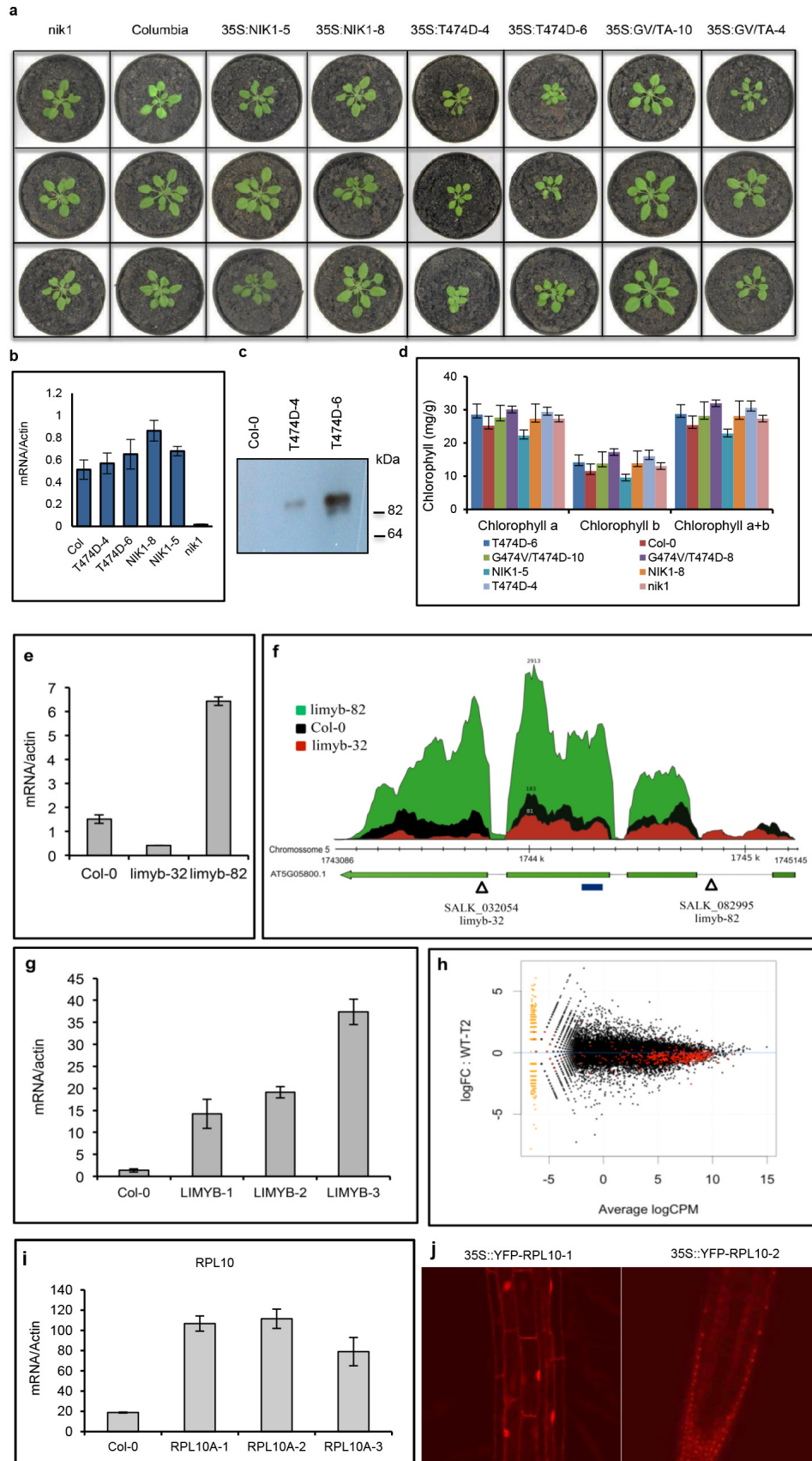
Extended Data Figure 1 | Characterization of the *Arabidopsis* transgenic lines. a, Phenotypes of wild-type (Col-0) and *nik1* plants transformed with NIK1 (NIK1-5 and NIK1-8), T474D (T474D-4 and T474D-6) or the double-mutant G4743V/T474A (inactive kinase, GV/TA-10 and GV/TA-4). Transgenic plants (R2 generation, $n = 15$) were grown in soil at 22 °C under short day conditions and photographed 2 weeks after planting. **b**, T474D transcript accumulation in transgenic lines (R2 generation). The expression of T474D or NIK1 in the leaves of independent transgenic lines was monitored by quantitative RT–PCR. Mean $\pm$ 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three independent experiments. **c**, Accumulation of T474D–GFP in transgenic lines. Total protein was extracted from the leaves of independent transgenic lines (as indicated), immunoprecipitated and immunoblotted with an anti-GFP antiserum. **d**, Chlorophyll content of transgenic lines. Total chlorophyll, chlorophyll *a* and chlorophyll *b* were determined in leaf sectors of the indicated transgenic lines. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of four independent experiments. **e**, Transcript accumulation of *LIMYB* in T-DNA insertion mutant lines. *LIMYB* expression was monitored by qRT–PCR of RNA prepared from Col-0, *limyb-32* (SALk_032054) and *limyb-82* (SALK_082995) plants. Gene expression was calculated using the $2^{-\Delta Ct}$ method, and actin was used as an endogenous control. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three independent experiments. **f**, Schematic representation of the *At5g05800* (*LIMYB*) locus in the chromosome 5 and RNA sequencing data. The *At5g05800* gene harbours three introns and four exons. Triangles show the positions of the T-DNA insertion in the *limyb-32* and *limyb-82* mutants, and the blue line indicates the position of the amplicon from **e**. The relative abundance of the mapped RNA hits in the *At5g05800* locus is shown in red in *limyb-32*, black in Col-0 and green in *limyb-82*. The accumulation of *LIMYB* transcripts was much lower in *limyb-32* and higher in *limyb-82* than in Col-0. Sequencing of the *limyb-32* and *limyb-82* transcripts revealed unprocessed intron sequences and premature stop codons that would have prevented the translation of a functional protein in these mutant lines. Therefore, *limyb-32* and *limyb-82* were confirmed as loss-of-function *limyb* mutant lines. **g**, *LIMYB* transcript accumulation in *LIMYB*-overexpressing lines (R2 generation). *LIMYB* expression in the leaves of independent transgenic lines was monitored by quantitative RT–PCR. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three independent experiments. **h**, General downregulation of translational machinery-related genes in LIMYB-1 seedlings. The 'MA' plots show the log of the ratio of the expression levels against log concentration, and each dot represents a gene. This plot visualizes the contrast of LIMYB-1 and Col-0 seedlings. The smear of points on the left side indicates those genes that were observed in only one group of replicated samples, and the red points denote ribosomal and protein synthesis-related genes. CPM, counts per million; FC, fold change; WT, wild type. **i**, *RPL10* transcript accumulation in *RPL10*-overexpressing lines (R2 generation). The expression of an NLS-containing *RPL10* transgene in the leaves of independent transgenic lines was monitored by quantitative RT–PCR. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three independent experiments. **j**, Nuclear localization of the NLS-containing YFP–RPL10 fusion in transgenic lines. Root tips from transgenic seedlings expressing the NLS-containing YFP–RPL10 fusion were directly examined under a laser confocal microscope. The figure shows representative confocal images from five independent biological replicates.

a

b

TAIR 10	log2 (Fold Change)	q-value (FDR)	Symbol	Description
AT2G14610.1	-2.0266	0.00166622880245705	PR1	Pathogenesis-related gene 1
AT4G11650.1	-1.4957	0.00338420995139227	OSM34	Osmotin 34
AT3G44480.1	-0.1045	0.763775753084199	RPP1, cog1	Disease resistance protein (TIR-NBS-LRR class)
AT4G16890.1	-0.0927	0.81956561493713	SNC1, BAL	Suppressor of NPR1
AT4G16950.1	-0.0397	0.923791385238997	RPP5	Disease resistance protein (TIR-NBS-LRR class)
AT4G33720.1	0.0	0.0	PR1	Pathogenesis-related 1 protein
AT5G17970.1	0.0	0.0		Disease resistance protein (TIR-NBS-LRR class)
AT5G40100.1	0.1313	1		Disease resistance protein (TIR-NBS-LRR class)
AT5G46490.1	0.2537	0.557766109000824		Disease resistance protein (TIR-NBS-LRR class)
AT5G67330.1	0.3322	0.497337654351076	NRAMP4	Natural resistance associated macrophage protein 4
AT3G55230.1	0.8647	0.293804289942712		Disease resistance-responsive (dirigent-like protein)
AT5G40020.1	1.1202	0.360825407950067		Pathogenesis-related thaumatin superfamily protein
AT1G63750.1	1.2229	0.209031413356128		Disease resistance protein (TIR-NBS-LRR class)
AT1G33560.1	1.3284	0.0253419904795418	ADR1	Disease resistance protein (CC-NBS-LRR class)
AT5G22690.1	1.8197	0.0106685101569183		Disease resistance protein (TIR-NBS-LRR class)
AT1G75040.1	2.8732	0.000722106289807284	PR-5	Pathogenesis-related gene 5

c

Extended Data Figure 2 | Expression of immune-system-related genes in T474D-overexpressing lines and in *LIMYB*-overexpressing lines. a–c, On the basis of our global comparison of EST sequences (Fig. 1a) and the role of NIK as an antiviral receptor, we asked whether the constitutive activation of NIK would elicit a defence response similar to that induced by geminivirus infection via the salicylic acid pathway or typical defence responses to virus. No significant gene enrichment was detected in the virus-induced gene silencing (GO:0009616) and viral defence response (GO:0051607) categories using the gene set enrichment analysis (GSEA) method (Supplementary Table 2). For the immune system category, gene enrichment was found in both up- and downregulated changes using the GSEA method. However, the typical markers of salicylic acid signalling, such as *PR1* and *SNC1*, were either non-differentially expressed or downregulated, and the expression of T474D did not enhance the salicylic acid level in the transgenic lines. Collectively, these results indicate that ectopic expression of T474D did not activate typical viral defences, such as salicylic acid signalling or gene silencing. **a, b,** Transcript accumulation of selected immune-system-related gene markers by RT–PCR (**a**) or RNA-sequencing in T474D-overexpressing lines (**b**). qRT–PCR of a representative sample confirmed an 80% match with the RNA-sequencing results. **c,** Transcript accumulation of the immune-system-related genes in LIMYB-overexpressing lines. The expression of the indicated genes in the leaves of independent transgenic lines was monitored by qRT–PCR. **a, c,** Mean ± 95% confidence intervals (*n* = 3) are shown, based on bootstrap resampling replicates of three independent experiments.
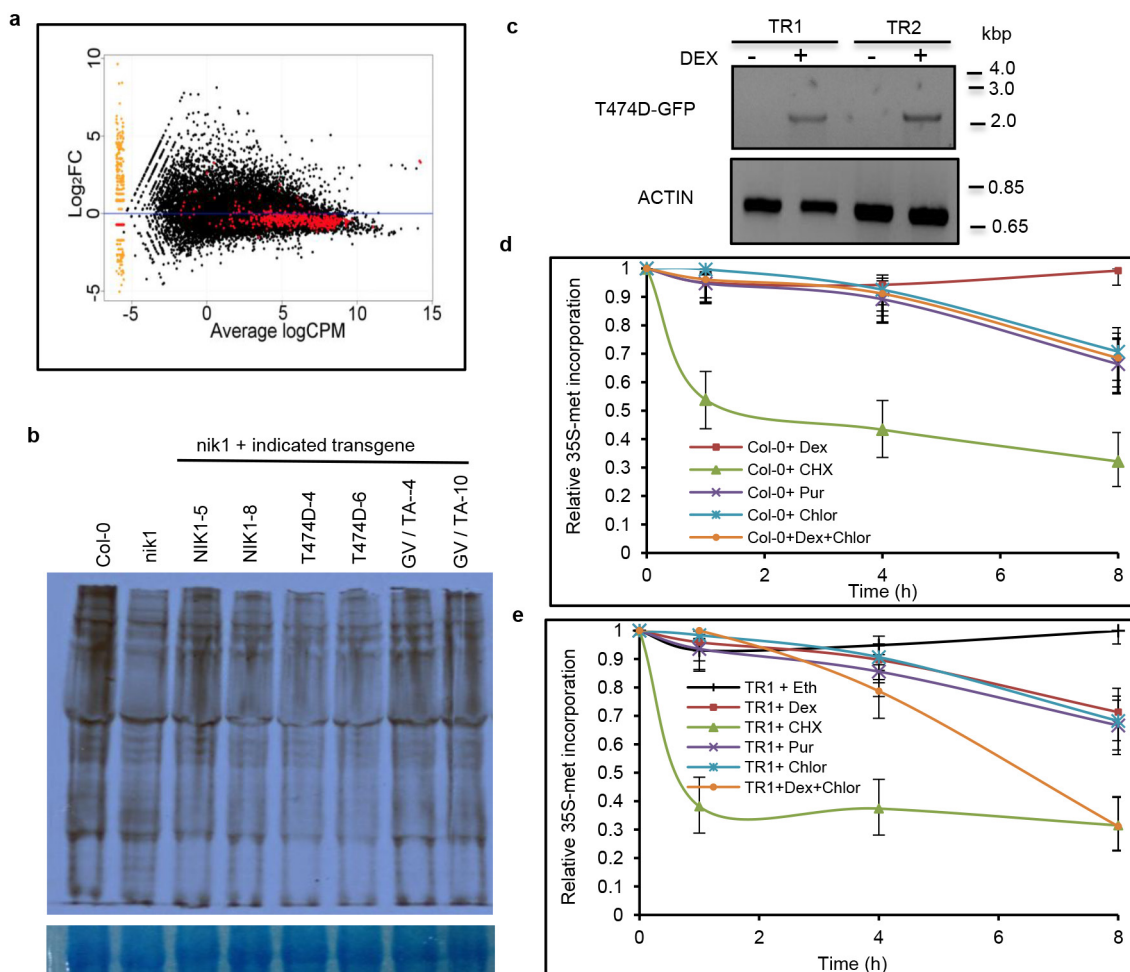
Extended Data Figure 3 | Ectopic expression of T474D-D downregulates translational-machinery-related genes and suppresses *de novo* protein synthesis. a, Representation of the translational-machinery-related genes differentially expressed in the T474D lines. The 'MA' plots show the log of the ratio of expression levels versus the log concentration, and each dot represents a gene. This plot represents the contrast between the T474D mock-inoculated lines and the Col-0 mock-inoculated lines. The red points denote ribosomal and protein synthesis-related genes (as shown in Supplementary Table 3). **b**, Downregulation of global translation by ectopic expression of T474D in *Arabidopsis*. The *in vivo* labelling of leaf proteins with [^{35}S]Met was performed in 20-day-old Col-0 plants and T474D transgenic lines. The total protein extracts were fractionated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), and the radioactive bands were quantified by densitometric analysis of the images obtained by autoradiography. The labelling percentage was normalized to the leaf chlorophyll content, and the protein loading on the gel was adjusted to the Coomassie-stained band of the large subunit of rubisco. **c**, Induction of T474D expression by dexamethasone (DEX) in *Arabidopsis* transgenic seedlings. The constitutive expression of T474D was associated with stunted growth in the transgenic lines (Extended Data Fig. 1a) and repression of global protein synthesis (Fig. 1b). These phenotypes precluded the use of an appropriate normalization method for protein loading in comparative gels of contrasting genotypes to estimate precisely the T474D-mediated protein synthesis inhibition in our assay. To overcome this limitation, we used a dexamethasone-inducible promoter to control the T474D expression in the transgenic lines. *Arabidopsis* seedlings independently transformed with a T474D–GFP fusion under the control of a dexamethasone-inducible promoter (TR1 and TR2) were treated with 30 μM dexamethasone for 8 h, and the induction of T474D–GFP expression was monitored by semi-quantitative RT–PCR. The dexamethasone-induced expression of T474D for 8 h led to a higher inhibition of *de novo* protein synthesis in the transgenic lines, as measured

by TCA-precipitable radioactivity, which could be normalized to total protein (TR1 and TR2; Fig. 1c). **d**, Inhibition of *de novo* protein synthesis by protein synthesis inhibitors in Col-0, untransformed lines. We also compared the T474D-mediated suppression of translation with known global translation inhibitors, such as the cytosolic protein synthesis inhibitors cycloheximide and puromycin and the chloroplast translation suppressor chloramphenicol. *Arabidopsis* seedlings (10 days old) were treated with 10 μM cycloheximide (Cyclo), 10 μM puromycin (Pur), 25 μM chloramphenicol (Chlor) or 30 μM dexamethasone (Dex) for the indicated periods of time, and then they were pulse labelled with L-[^{35}S]Met for 60 min. Lysates of treated cells were measured by liquid scintillation counting and normalized to total protein. The relative [^{35}S]Met incorporation was normalized to wild-type (WT = 1) control without treatment. Means ± 95% confidence intervals (*n* = 3) based on bootstrap resampling replicates of three independent experiments are shown. **e**, Inhibition of *de novo* protein synthesis by inducible expression of T474D. Seedlings (10 days old) from the TR1 transgenic line were treated with dexamethasone and the protein synthesis inhibitors for the times as indicated in the figure, and then they were pulse labelled with L-[^{35}S]Met for 60 min. Lysates were processed as described in **d**. Means ± 95% confidence interval (*n* = 3) based on bootstrap resampling replicates of three independent experiments are shown. Cycloheximide was the most effective inhibitor of translation in both the wild-type and T474D-expressing lines. T474D expression inhibited global translation to the same extent as puromycin and caused a further inhibition in the level of chloramphenicol translational inhibition in a combined treatment. The increase in translational inhibition by combining T474D expression and chloramphenicol treatment may indicate that T474D inhibits cytosolic protein synthesis, which is consistent with the T474D-mediated downregulation of components of the cytosolic translational machinery (Supplementary Table 3).

a

b

c

Relative RNA content in NPS and PS complexes

	NPS	PS	Total
TR1	0.45 ± 0.03	0.55 ± 0.02	100%
TR1+ Dex	0.39 ± 0.02	0.48 ± 0.02	87%

d

e

f

g

h
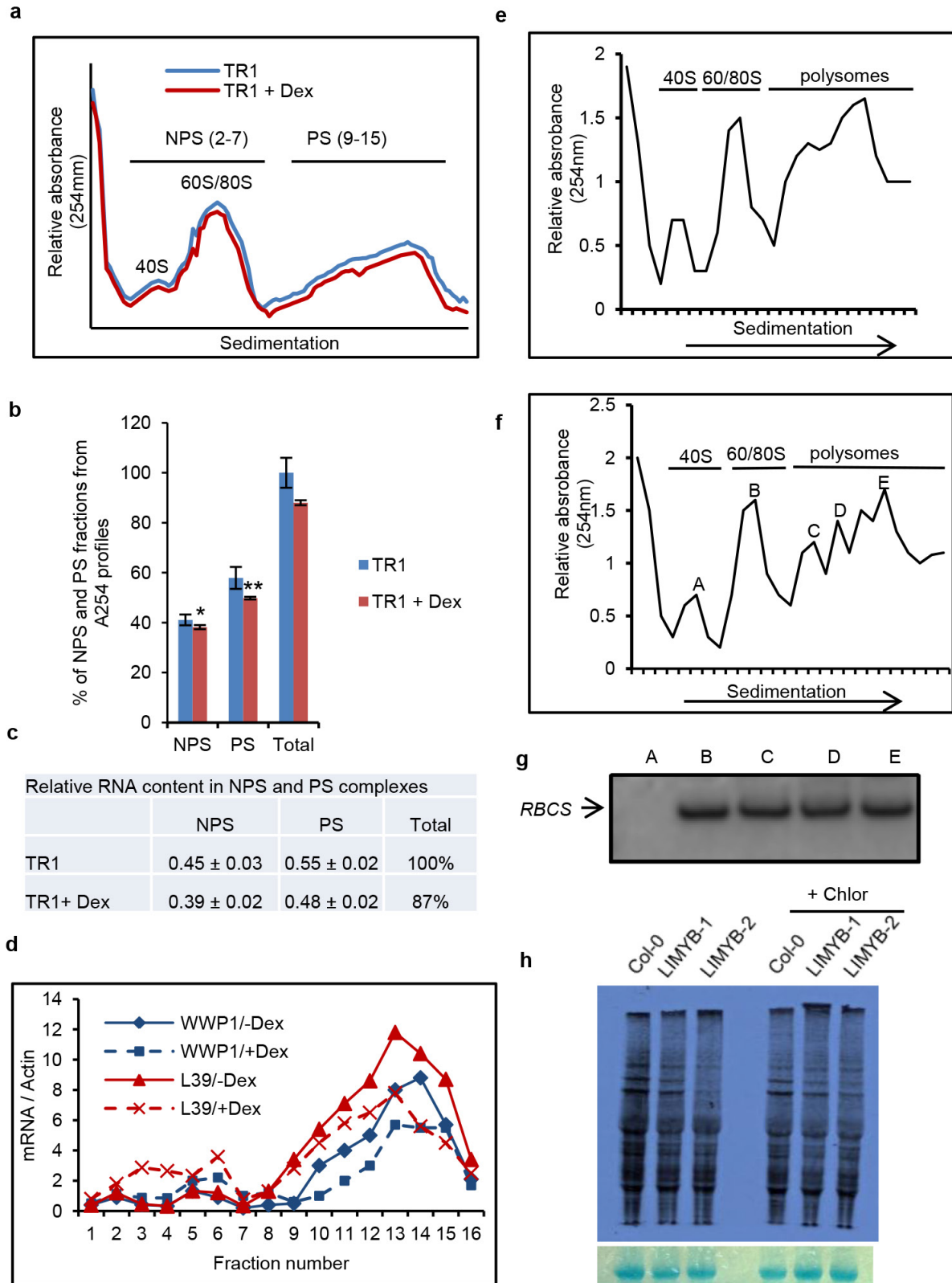
Extended Data Figure 4 | Isolation of PS fractions from *Arabidopsis* seedlings and LIMYB-mediated inhibition of protein synthesis.
a, Ultraviolet absorbance profiles of the sucrose gradient used for RNA fractionation of dexamethasone (Dex)-inducible T474D transgenic lines. Sixteen fractions of 310 µl were collected. NPS RNA includes complexes ≤ 80S that fractionated in the top half of the gradient (fractions 2–7) and PS (polysome) represents complexes fractionated in the bottom of the gradient (fractions 9–15). **b**, Distribution (%) of NPS and PS fractions on polysome density gradients. The percentage of NPS or PS was calculated by integrating the areas under the corresponding peaks in the $A_{254\,nm}$ profile delimited by a gradient baseline absorbance (buffer density). Values are the average ± standard deviation (s.d.) of three biological replicates. NPS and PS fractions from T474D + Dex samples were significantly different from the corresponding fractions of the T474D samples by the *t*-test (greater). NPS, *P value = 0.03725; PS, **P value = 0.009137 (*t*-test; greater). **c**, Relative RNA content in NPS and PS complexes. RNA was precipitated from NPS and PS density regions of sucrose gradients (as in **a**) and quantified. Relative NPS RNA and PS RNA contents from T474D and T474D + Dex samples were calculated in relation to the total NPS + PS content from T474D. Values for the relative NPS and PS RNA content are the average ± s.d. of three biological replicates and they were significantly different between the samples (*P* < 0.01,

t-test). **d**, Distribution of specific mRNAs in the PS gradient fractions from extracts prepared from T474D seedlings treated (or not) with dexamethasone. The RNA on each fraction was reverse transcribed and aliquots amplified with specific primers for the indicated genes by qPCR. **e**, Ultraviolet absorbance profiles of the sucrose gradient used for the RNA fractionation of Col-0 seedlings. PSs from 15-day-old Col-0 seedlings were fractionated on a sucrose gradient, and the fractions were manually collected. **f**, Ultraviolet absorbance profiles of the sucrose gradient used for RNA fractionation from T474D seedlings. PSs from 15-day-old T474D-overexpressing seedlings were fractionated on a sucrose gradient, and the fractions were manually collected. **g**, Levels of the small subunit of rubisco (*RBCS*) mRNA per fraction. The levels of mRNA of *RBCS* were examined by northern blotting. This control was used to ensure the quality and distribution of a specific mRNA. **h**, Overexpression of LIMYB suppresses cytosolic translation. *In vivo* labelling of leaf proteins with [^{35}S]Met was performed in Col-0 and *LIMYB-1* transgenic seedlings in the presence and absence of chloramphenicol treatment. The total protein extracts were fractionated by SDS–PAGE, and the radioactive bands were quantified by densitometric analysis of the images obtained by autoradiography. The labelling percentage was normalized to the leaf chlorophyll content, and protein loading is shown by Coomassie staining of the radioactive gel.

a

c

T474D infected – Col-0 infected (10 days)			
Viral Component	log2FC	q-value	NCBI Locus ID
A	0.9139909	1.381925e-01*	CLU65529
B	1.2348528	1.257184e-08	CLU65530
T474D infected – WT infected (21 days)			
Viral Component	log2FC	q-value	NCBI Locus ID
A	-4.534558	7.640602e-126	CLU65529
B	-3.682993	1.129406e-112	CLU65530

*Non significant according to Negative Binomial Exact Test (edgeR) by q-value (FDR) < 0.05

b

d

Extended Data Figure 5 | Ectopic expression of T474D and LIMYB confers tolerance to begomovirus infection. a, Delayed onset of infection in the T474D-4- and T474D-6-overexpressing lines. Ecotype Col-0 plants, as well as the T474D-4- and T474D-6-overexpressing lines, and the *NIK1-8* and G473V/T474A-overexpressing lines were infected with CaLCuV DNA by the biolistic method. The progression of the infection was monitored by PCR detection of viral DNA in the systemic leaves of the inoculated plants. The values represent the percentages of systemically infected plants at different dpi. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of four independent experiments. **b**, Upon symptom development, the T474D-induced transcriptome diverges from the infected Col-0 transcriptome. The mock-inoculated T474D-overexpressing lines exhibited a constitutively infected wild-type transcriptome at 10 dpi (Fig. 1a). Nevertheless, these T474D transgenic lines did not phenocopy the infected wild-type plants because they did not develop symptoms of viral infection. In fact, the wild-type plants displayed typical symptoms of CaLCuV infection at 21 dpi, such as leaf distortion, stunting with epinasty and chlorosis (Fig. 1e). The symptoms in the T474D-expressing lines at 21 dpi, however, were greatly attenuated, with no visible leaf distortion or chlorosis. To examine these phenotypes, we performed a Ward hierarchical clustering of the gene expression data (normalized by the trimmed mean of *M*-values (TMM) normalization method) from the *Arabidopsis* infection experiments at 21 dpi. The TMM normalization method assumes that the majority of genes are not differentially expressed, and it adjusts genes with larger read counts and lower variance on the logarithmic scale. The dendrogram provides two types of *P* values: AU (black) and BP (grey). The AU *P* value comes from multiscale bootstrap resampling, while the BP value represents normal bootstrap resampling. These *P* values were calculated by multiscale bootstrap resampling using the R-cran package pvclust with a cut-off of 0.05. These *P* values show the significance of the proximity of each gene expression experiment profile. The cluster analysis at 21 dpi indicated that when symptoms had developed in the infected Col-0 leaves, the T474D-induced transcriptome diverged from the infected Col-0 transcriptome. The mock T474D transcriptome formed a unique clade, while the infected T474D transcriptome was more closely related to the mock-inoculated Col-0 transcriptome. **c**, Reduced viral transcript accumulation in T474D-overexpressing lines at 21 dpi. RNA-sequencing data of viral gene transcripts in the systemic leaves of infected wild-type and T474D-overexpressing plants at 10 dpi and 21 dpi. **d**, The onset of infection is delayed in *LIMYB*-overexpressing lines. Ecotype Col-0 plants and *LIMYB*-overexpressing and *limyb* mutant lines were infected with CaLCuV DNA using the biolistic method. The progression of the infection was monitored by the PCR detection of viral DNA in the systemic leaves of the inoculated plants. The values represent the percentages of systemically infected plants at different dpi. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of four independent experiments.
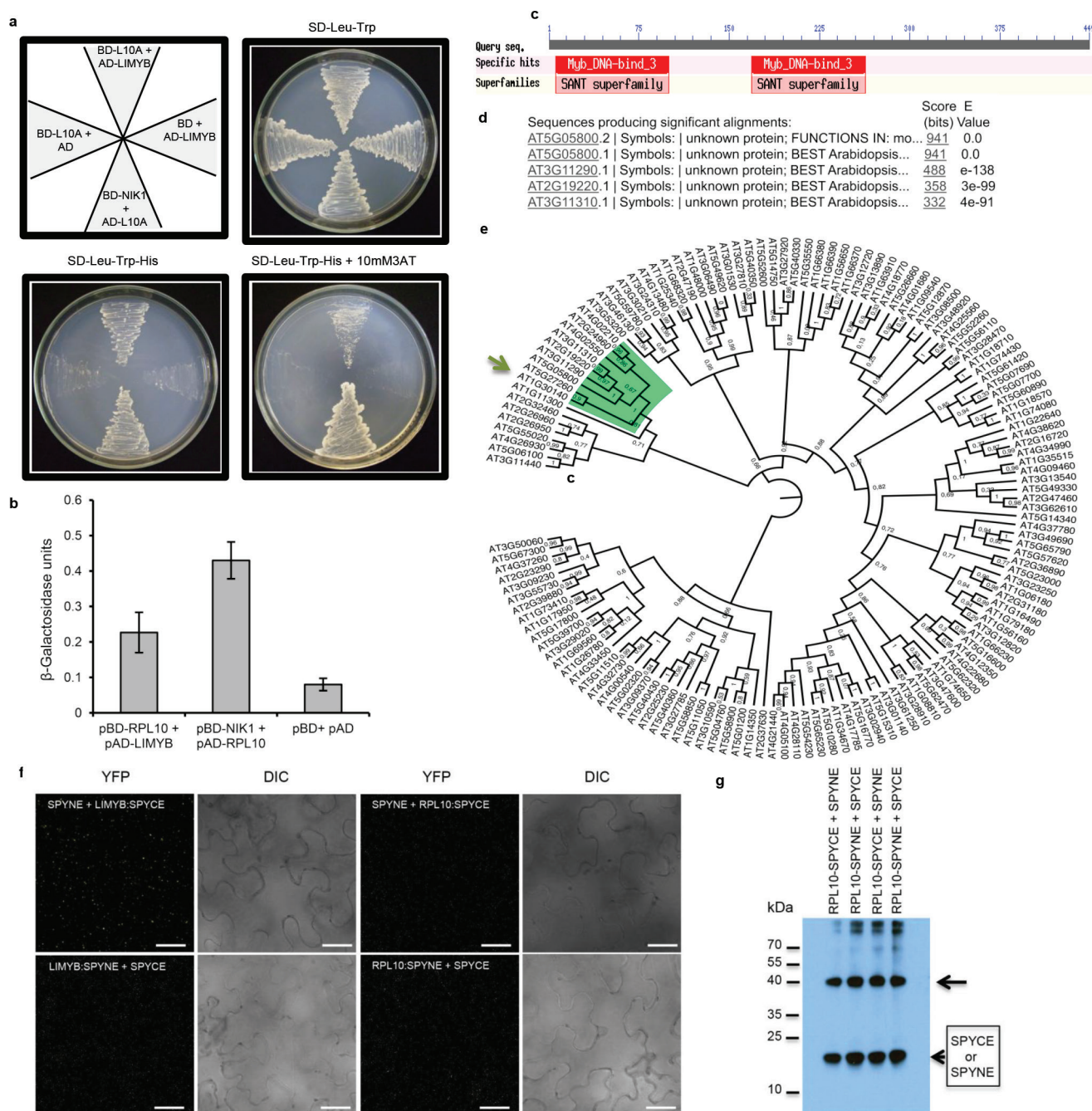
Extended Data Figure 6 | LIMYB, which belongs to the MYB domain-containing superfamily, interacts with RPL10 in the yeast two-hybrid system. a, LIMYB and RPL10 interact in yeast. LIMYB was expressed in yeast as a GAL4 activation domain (AD) fusion (pAD–LIMYB), and RPL10 was expressed as a GAL4 binding domain (BD) fusion (pBD–RPL10). The interactions between the tested proteins were examined by monitoring His prototrophy. **b,** The interactions were further confirmed by measuring the expression activity of the β-galactosidase reporter enzyme for the second reporter gene, β-galactosidase. The interaction between pAD–RPL10 and pBD–NIK1 was monitored as a positive control. Means ± 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three technical replicates are shown. **c,** LIMYB harbours two MYB domains. The position of these MYB domains is indicated in the schematic representation of the LIMYB primary structure. **d,** Sequence identity of the closest related *LIMYB* (*At5g05800*) homologues. **e,** Dendrogram of MYB domain-containing proteins from *Arabidopsis*. The MYB family sequences were retrieved from the Agris database (http://arabidopsis.med.ohio-state.edu). The alignment was performed by Maft aligner software using full-length sequences, and the tree

was built by Fasttree software (the bootstrap values are indicated close to the branch divisions). The arrow indicates *LIMYB* (*At5g05800*). **f, g,** The negative controls used in the BiFC analysis. **f,** Confocal fluorescent image of SPYNE + LIMYB:SPYCE, LIMYB:SPYNE + SPYCE, SPYNE + RPL10:SPYCE and RPL10:SPYNE + SPYCE, as indicated in the figure. We used a BiFC assay to determine whether RPL10 and LIMYB interact in the nuclei of plant cells. The formation of a RPL10–LIMYB complex occurred in the nuclei of transfected cells independent of the orientation of the LIMYB or RPL10 fusions (amino terminus or carboxy terminus of YFP; Fig. 2a), and the reconstituted fluorescent signal was much higher than that of the background (control panels with combinations of the protein fusions with empty vectors). The figure displays representative samples from three independent biological repeats. Scale bars, 20 μm. **g,** The C-terminal (SPYCE) and N-terminal region (SPYNE) of YFP accumulates detectably in co-transfected leaves. Total protein extracts from leaves co-transfected with the indicated constructs were immunoblotted with anti-YFP serum. The arrow indicates the position of RPL10–SPYCE and RPL10–SPYNE fusions, and arrowheads indicate the positions of the C-terminal (SYYCE) and N-terminal (SPYNE) regions of YFP.
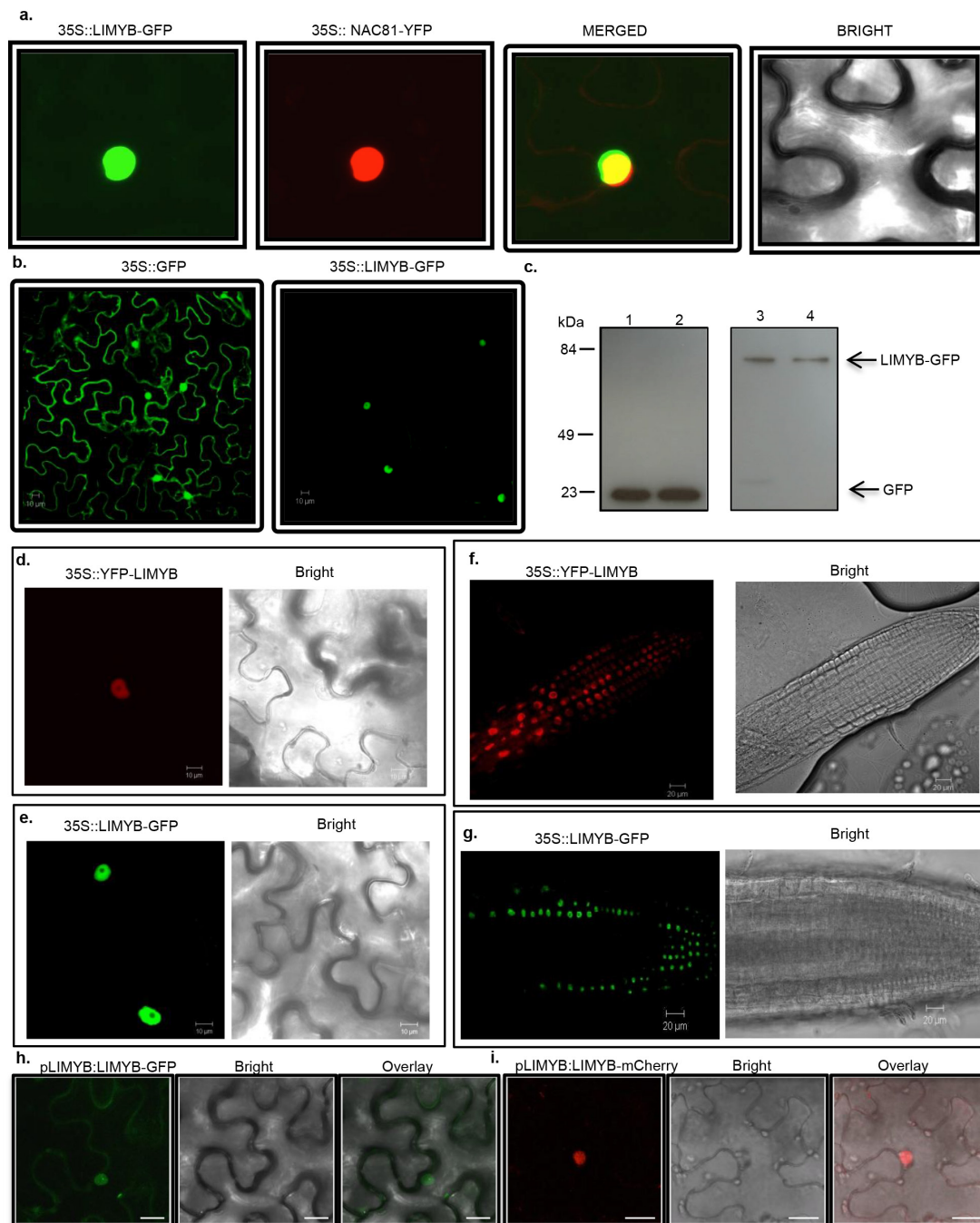
Extended Data Figure 7 | Nuclear localization of LIMYB. a, Colocalization of LIMYB with the nuclear marker gene *Glycine max* (*Gm*)*NAC81*. *N. benthamiana* leaves were co-infiltrated with *A. tumefaciens* carrying a 35S::LIMYB–GFP construct and a 35S::YFP–GmNAC81 construct. Forty-eight hours after infiltration, the subcellular localizations of the fluorescent fusion proteins were examined by confocal microscopy. The figure shows representative confocal images from two independent experiments. **b,** Confocal fluorescence image of transiently expressed GFP (left) or LIMYB–GFP (right) in epidermal cells of tobacco leaves. Scale bars, 10 μm. The figure shows representative confocal images from two independent experiments. **c,** Immunoblotting of transiently expressed LIMYB–GFP in epidermal cells of tobacco leaves. Total protein was extracted from agro-infiltrated *N. benthamiana* leaves containing the 35S::GFP (left lanes) or 35S::LIMYB–GFP (right lanes) constructs and immunoblotted with an anti-GFP monoclonal antibody to examine the stability of the fusion protein. The positions of molecular mass are shown in kDa. **d,** Confocal fluorescence image of transiently expressed GFP–LIMYB in epidermal cells of tobacco leaves. Scale bars, 10 μm. The figure shows representative confocal images from four independent

experiments. **e,** Confocal fluorescence image of transiently expressed LIMYB–GFP in epidermal cells of tobacco leaves. Scale bars, 10 μm. The figure shows representative confocal images from four independent experiments. **f, g,** Confocal fluorescence image of root cells stably transformed with YFP–LIMYB or LIMYB–GFP. Root tips from transgenic seedlings expressing YFP–LIMYB (**f**) or LIMYB–GFP (**g**) were directly examined under a laser confocal microscope. Scale bars, 20 μm. The figures show representative confocal images from three biological replicas. Neither the fusion of YFP to the LIMYB N terminus nor GFP to its C terminus altered the nuclear localization of LIMYB in either agro-inoculated *N. tabacum* leaves or stably transformed *Arabidopsis* roots. **h, i,** Confocal fluorescent image of LIMYB fused to GFP or mCherry under the control of its own promoter. The figures show representative confocal images from two independent experiments. The fluorescence was also concentrated in the nucleus of transfected cells by expression of LIMYB–GFP or LIMYB–mCherry fusions under the control of the LIMYB endogenous promoter. Scale bars, 20 μm. Collectively, these results indicate that LIMYB was localized in the nucleus.
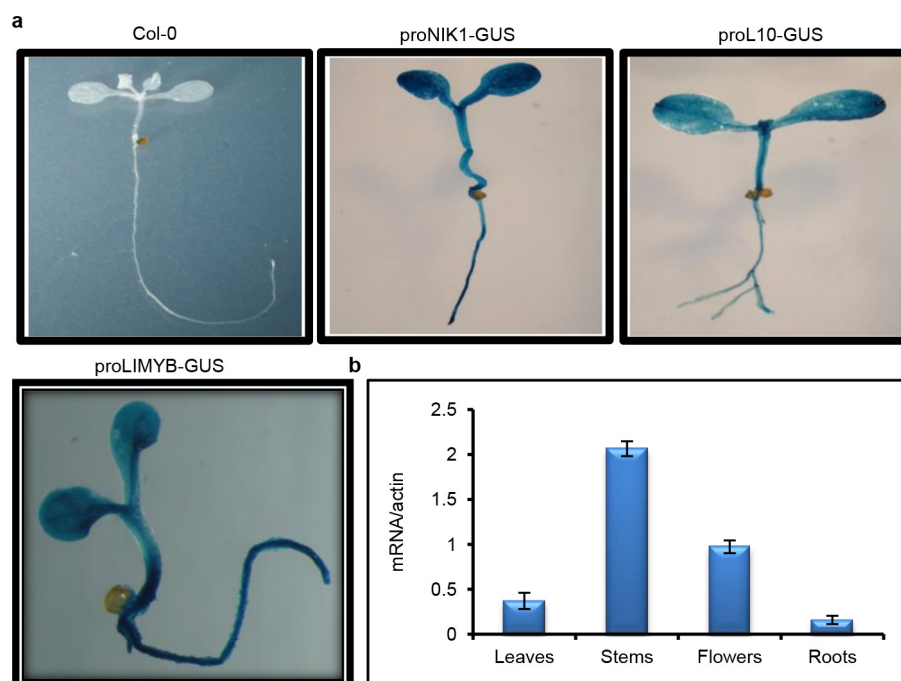
Extended Data Figure 8 | *LIMYB*, *RPL10* and *NIK1* display overlapping expression profiles. **a**, pLIMYB::GUS, pRPL10::GUS and pNIK1::GUS are ubiquitously expressed in seedling tissues. GUS reporter gene expression was histochemically monitored in 2-week-old seedling leaves and roots from transgenic lines harbouring a β-glucuronidase (*GUS*) reporter gene expressed from the LIMYB, RPL10 and NIK1 promoters. The figure shows representative GUS staining images of three seedlings per genotype. All three genes were ubiquitously expressed in all seedling tissues. **b**, Expression analysis of LIMYB in various plant organs. *LIMYB* expression was monitored by qRT–PCR of RNA prepared from leaves, roots, stems or flowers of Col-0 plants. Gene expression was calculated using the $2^{-\Delta Ct}$ method, and actin was used as an endogenous control. Error bars, 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three independent experiments. LIMYB was also expressed in the leaves, roots, stems and flowers, indicating that *LIMYB*, *RPL10* and *NIK1* are co-expressed in several organs.
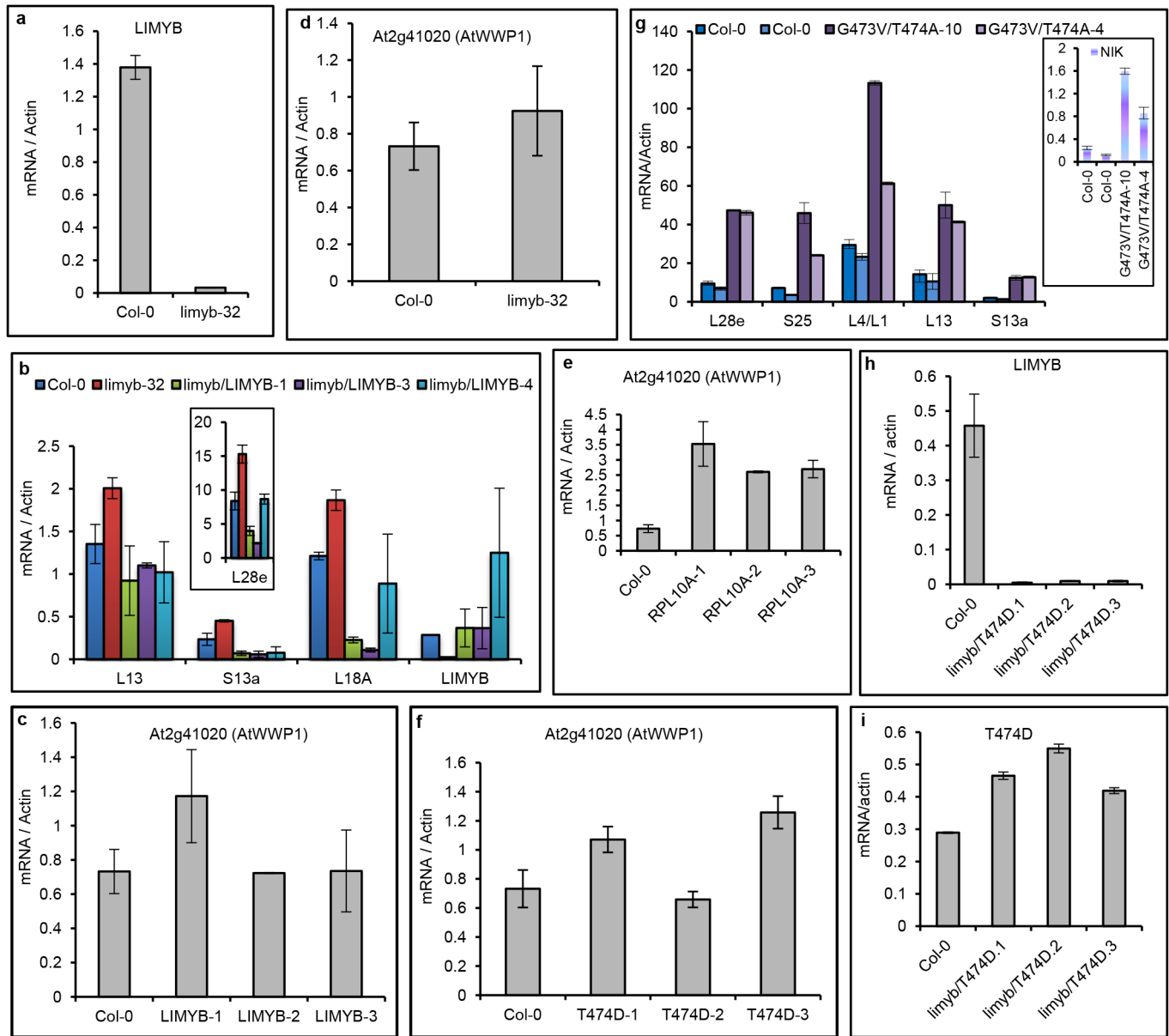
Extended Data Figure 9 | Controls for the regulation of RP gene expression experiments. a–i, Expression of the indicated genes in the leaves of independent transgenic lines was monitored by qRT–PCR. a, *LIMYB* expression in the *limyb-32* mutant was examined. b, *LIMYB* expression in *limyb-32* mutant restores wild-type expression of the RP genes. Expression of the *S13a*, *L18A* and *L28e* genes was monitored in three independently transformed *limyb-32* knockout plants with the *LIMYB* gene. c–f, Expression of the unrelated gene *AtWWP1* was monitored as a negative control in three independently transformed *LIMYB*-, *RPL10*- and T474D-overexpressing lines in addition to the *limyb-32* mutant. g, The double-mutant inactive kinase, G4743V/T474A, does not downregulate the RP genes. The transcript accumulation of the indicated RP genes was quantified by qRT–PCR in two independently transformed *nik1* knockout lines expressing the G4743V/T474A double mutant. h, i, Expression of *LIMYB* (h) and the transgene T474D (i) was monitored in the *limyb-32* lines, which were transformed with T474D. a–i, Means ± 95% confidence intervals (*n* = 3) based on bootstrap resampling replicates of three independent experiments are shown.
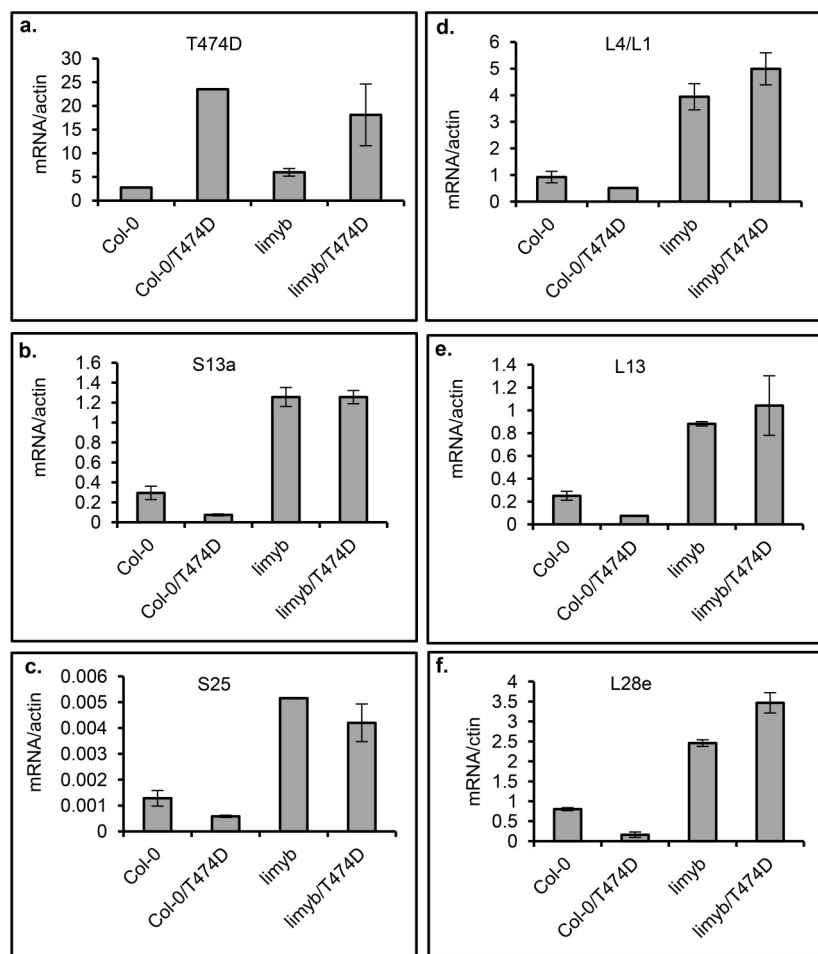
Extended Data Figure 10 | T474D requires LIMYB function to mediate the downregulation of translational machinery-related genes. a–f, A 35S::T474D–GFP construct was electroporated into protoplasts prepared from Col-0 and *limyb-32* seedlings, and the expression of the indicated RP genes was monitored by qRT–PCR of RNA prepared from untransfected and transfected protoplasts. Gene expression was calculated using the $2^{-\Delta Ct}$ method, and actin and ubiquitin were used as an endogenous control. Means ± 95% confidence intervals ($n = 3$) based on bootstrap resampling replicates of three independent experiments are shown.

LETTER

A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria

Alassane Mbengue[1,2]*, Souvik Bhattacharjee[1,2]*, Trupti Pandharkar[1,2], Haining Liu[1,3], Guillermina Estiu[1,3]‡, Robert V. Stahelin[1,3,4], Shahir S. Rizk[1,2], Dieudonne L. Njimoh[1,2,5], Yana Ryan[1,2], Kesinee Chotivanich[6], Chea Nguon[7], Mehdi Ghorbal[8], Jose-Juan Lopez-Rubio[8], Michael Pfrender[2], Scott Emrich[9], Narla Mohandas[10], Arjen M. Dondorp[6,11], Olaf Wiest[1,3,12] & Kasturi Haldar[1,2]

Artemisinins are the cornerstone of anti-malarial drugs[1]. Emergence and spread of resistance to them[2–4] raises risk of wiping out recent gains achieved in reducing worldwide malaria burden and threatens future malaria control and elimination on a global level. Genome-wide association studies (GWAS) have revealed parasite genetic loci associated with artemisinin resistance[5–10]. However, there is no consensus on biochemical targets of artemisinin. Whether and how these targets interact with genes identified by GWAS, remains unknown. Here we provide biochemical and cellular evidence that artemisinins are potent inhibitors of *Plasmodium falciparum* phosphatidylinositol-3-kinase (PfPI3K), revealing an unexpected mechanism of action. In resistant clinical strains, increased PfPI3K was associated with the C580Y mutation in *P. falciparum* Kelch13 (PfKelch13), a primary marker of artemisinin resistance. Polyubiquitination of PfPI3K and its binding to PfKelch13 were reduced by the PfKelch13 mutation, which limited proteolysis of PfPI3K and thus increased levels of the kinase, as well as its lipid product phosphatidylinositol-3-phosphate (PI3P). We find PI3P levels to be predictive of artemisinin resistance in both clinical and engineered laboratory parasites as well as across non-isogenic strains. Elevated PI3P induced artemisinin resistance in absence of PfKelch13 mutations, but remained responsive to regulation by PfKelch13. Evidence is presented for PI3P-dependent signalling in which transgenic expression of an additional kinase confers resistance. Together these data present PI3P as the key mediator of artemisinin resistance and the sole PfPI3K as an important target for malaria elimination.

Our previous work identified an important role for PI3P in protein export from the *P. falciparum* endoplasmic reticulum (ER) to the erythrocyte, at the early 'ring' stage of blood infection[11]. Consequently, a secretory reporter that binds PI3P remains in the ring ER but in absence of PI3P undergoes default secretion to the parasitophorous vacuole (PV). This yielded a cell-based screen for drugs that inhibit PI3P production (Fig. 1a). We were particularly interested in artemisinins because clinical resistance to them develops at the early ring stage[3]. Low nanomolar concentrations of dihydroartemisinin (DHA), the active form of all artemisinins block production of PI3P (Fig. 1a). This effect is fast acting (within 30 min), reversed by washing out the drug and without effect on subsequent parasite growth (Extended Data Fig. 1a). Wortmannin or LY294002, active against the sole parasite PfPI3K[12,13], but not the inactive LY303511 blocked

PI3P production. Artemisinin and artesunate were also inhibitory (Extended Data Fig. 1b, c), but deoxyartemisinin, anti-folates and aminoquinolines had no effect (Fig. 1a and Extended Data Fig. 1b–e). Biochemical analyses confirmed that DHA reduced mass PI3P levels and drug washout restored PI3P levels (Fig. 1b). Quantitative inhibition of immunopurified PfPI3K was achieved by 4 nM DHA but not by deoxyartemisinin (Fig. 1c). DHA at 10 μM failed to significantly inhibit 46 mammalian kinases, including its closest human orthologue VPS34 (a class III kinase; Fig. 1d and Extended Data Table 1) strongly supporting the conclusion that DHA is not a promiscuous kinase inhibitor.

In the absence of a crystal structure of PfPI3K, a computational model can provide structural hypotheses to help understand experimental results. We used homology modelling of PfPI3K and 20 ns molecular dynamics (MD) simulations[14] to study the binding of DHA and several structural analogues (Fig. 1e–g, Extended Data Figs 2 and 3a and Supplementary Data 1). The model suggests polar contacts of DHA with the D1889 hydroxyl and the Y1915 lactol oxygen of PfPI3K at the binding site (Fig. 1f) as well as an excellent shape complementarity of DHA with PfPI3K at its hydrophobic region (Fig. 1g). This is in agreement with the rapid inhibition of PfPI3K by DHA at nanomolar concentrations. MD simulation also rationalize the experimentally observed lack of inhibition of VPS34, mammalian class I or class II PI3 kinases by DHA (Extended Data Fig. 3b–d). Artesunate (at high 10 μM concentrations) has been reported to inhibit activities of the PI3K–AKT pathway in mammalian systems[15–17]. However, whether this was due to direct inhibition of PI3K or AKT, the DHA metabolite or other indirect effects (such as high concentrations that far exceed inhibitory concentrations for malaria parasites) was not studied.

Together, the data in Fig. 1 suggest that DHA specifically targets PfPI3K. Yet GWAS studies suggest that >1,000 genes (including PfPI3K) show clinical resistance to artemisinins[6]; moreover PfPI3K polymorphisms are not detected in all resistant strains. Rather, there is selective pressure in regions of chromosome 13 (refs 5, 9) and in particular *pfkelch13* (refs 7, 8, 10) but the mechanism is unknown. The mammalian orthologue of PfKelch13 functions as a substrate adaptor for an E3 ubiquitin ligase[18]. The putative substrate binding domain of PfKelch13 shows characteristic 'Kelch' propeller domain[7,18] (Fig. 2a) mutations in which associate with artemisinin resistance[7,8,10]. We hypothesized that mutations may decrease affinity for a protein

Figure 1 | Dihydroartemisinin targets PfPI3K.
a, SS-EEA1[WT]-mCherry detects ring PI3P in punctate (ER) domains[11]. Mutant SS-EEA1[R1374A]-mCherry secretes to the PV[11] (second row). Treatment with 4 nM DHA redistributes SS-EEA1[WT]-mCherry to the PV. Washout restores ER PI3P. Treatment with 4 nM deoxyartemisinin had no effect. Blue, nucleus; scale bar, 5 μm; P, parasite; E, Erythrocyte. Mean (s.d.) with three experimental replicates with image analysis from 400 optical sections. **b–d**, Effects of DHA on PI3P mass (**b**); immunopurified PfPI3K (raw data in Supplementary Data 2) (**c**); and mammalian PI3 kinases (**d**). Mean from three experimental replicates (each with triplicate data points). For **b**, s.d. <3; **c**, upper graph, s.d. <1.5; lower graph s.d. <5; **d**, s.d. <0.5. **e**, Overlay of the model of PfPI3K (cyan) and human class III PI3KVPS34 (grey, PDB code 3IHY) with active site marked (asterisk). **f**, DHA in PfPI3K model (cyan) binding site. **g**, Surface representation of **f**. Additional details in Extended Data Figs 1–3.

substrate, thereby increasing its steady state levels by reducing ubiquitination and (proteosomal) degradation.

We tested whether changes in PfPI3K levels were associated with PfKelch13 mutations in clinical and engineered laboratory strains. Analysis of clinical isolates from Cambodia[19] (also see Methods) suggested a ~1.5-fold to twofold increase in PfPI3K levels in resistant versus sensitive strains (Fig. 2b). One sensitive clinical strain (ANL-7) contained the resistance mutation R539T: genome sequencing revealed it was not clonal suggesting contamination with a resistant parasite strain (data not shown). The PfKelch13 mutation C580Y was not present in clinical sensitive strains but detected in two of three resistant strains. C580Y was also recognized to be the most prevalent mutation in resistant populations[7,8] and thus investigated further in two distinct laboratory strains. We used parasites with chromosomally inserted PfKelch13[C580Y] in the *P. falciparum* NF54 (ref. 20) (Extended Data Fig. 4a). Additionally, we expressed a HA-tagged form of PfKelch13[C580Y] in a second strain 3D7 (Extended Data Fig. 4b and Extended Data Table 2). Both mutated strains showed twofold to threefold increase in levels of PfPI3K relative to their PfKelch13[WT] counterparts (Fig. 2c, d) without changes in levels of PfKelch13 (Extended Data Fig. 4a, c).

Further analysis revealed that immunopurified (IP) PfPI3K was bound to PfKelch13[WT] and this was reduced by the C580Y mutation (Fig. 2e). K48-linked and atypical ubiquitination, both signals for proteosomal degradation were prominent in PfPI3K and concomitantly

reduced by PfKelch13[C580Y]. K63-linked ubiquitination[21] was not detected suggesting ubiquitination drives degradation rather than a change in cellular localization and endocytic trafficking and PfPI3K continued to be cytoplasmically distributed as seen in absence of mutation (Extended Data Fig. 1f). Fragmentation detected in PfPI3K was diminished in presence of the C580Y mutation. Cleavage of the kinase began even before it was fully released from PfKelch13[WT] and confirmed that lack of effect of PfKelch13[C580Y] was due to its failure to bind PfPI3K (Fig. 2f). Notably, Kelch adapters for E3 ligases are substrate specific and therefore our data well support the model in Fig. 2g. As PfPI3K is an early ring-stage target of artemisinins, elevation of the kinase and/or its products may provide a mechanism of artemisinin resistance at this stage. In an isogenic background, PfKelch13 mutation linked to resistance may increase levels of PfPI3K (Fig. 2c). Nonetheless equal amounts of PfPI3K across two distinct non-isogenic strains may not show equivalent activity for a variety of reasons, including different levels of precursor lipid substrate pools. We therefore examined whether PfPI3K activity, as measured by PI3P production provided a quantitative estimation of resistance across non-isogenic strains. PI3P can only be produced by (the sole) PfPI3K[12]. PI3,4P2 and PI3,4,5P3 (derived from PI3P) are both detected as minor fractions of the total cellular PI3P and under 1% in rings[12] the stage of clinical artemisinin resistance. Thus ring parasite PI3P levels directly reflect the activity of PfPI3K. Figure 3a indicated there was linear correlation between the levels of PI3P and resistance

Figure 2 | Proteostatic regulation of PfPI3K by PfKelch13. a, PfKelch13 'Kelch' propeller domain. **b,** PfPI3K, PfKelch13 polymorphisms, PfPI3K levels in clinical (ANL) resistant (2, 4, and 9) and sensitive (1, 7) strains and laboratory strain 3D7; loading control PfFKBP. ANL-7, contaminated with resistance mutation R539T. **c, d,** Western blots show PfPI3K in NF54-PfKelch13^{C580Y} and NF54-PfKelch13^{WT}; 3D7-PfKelch13^{WT}-HA and 3D7- PfKelch13^{WT}-HA; loading controls, BiP, PfFKBP. **e,** Top left, PfKelch13 (arrow) binding to PfPI3K is reduced by PfKelch13^{C580Y}. In the PfKelch13^{WT} background, IP PfPI3K displays K48-ubiquitination (top middle, arrow), atypical poly-ubiquitination (bottom left, asterisk) and degradation (bottom right), all reduced by PfKelch13^{C580Y}. K63-ubiquitination was not observed (top right). IP lysate input PfFKBP. **f,** PfKelch13 but not PfKelch13^{C580Y} binds PfPI3K (arrow). Loading control, PfFKBP. Molecular weights in kDa. In **b–f**, data are representative of three experimental replicates. **g,** A model for Fig. 2. Additional details in Extended Data Fig. 4. Raw data for **b–f** is in Supplementary Data 2.
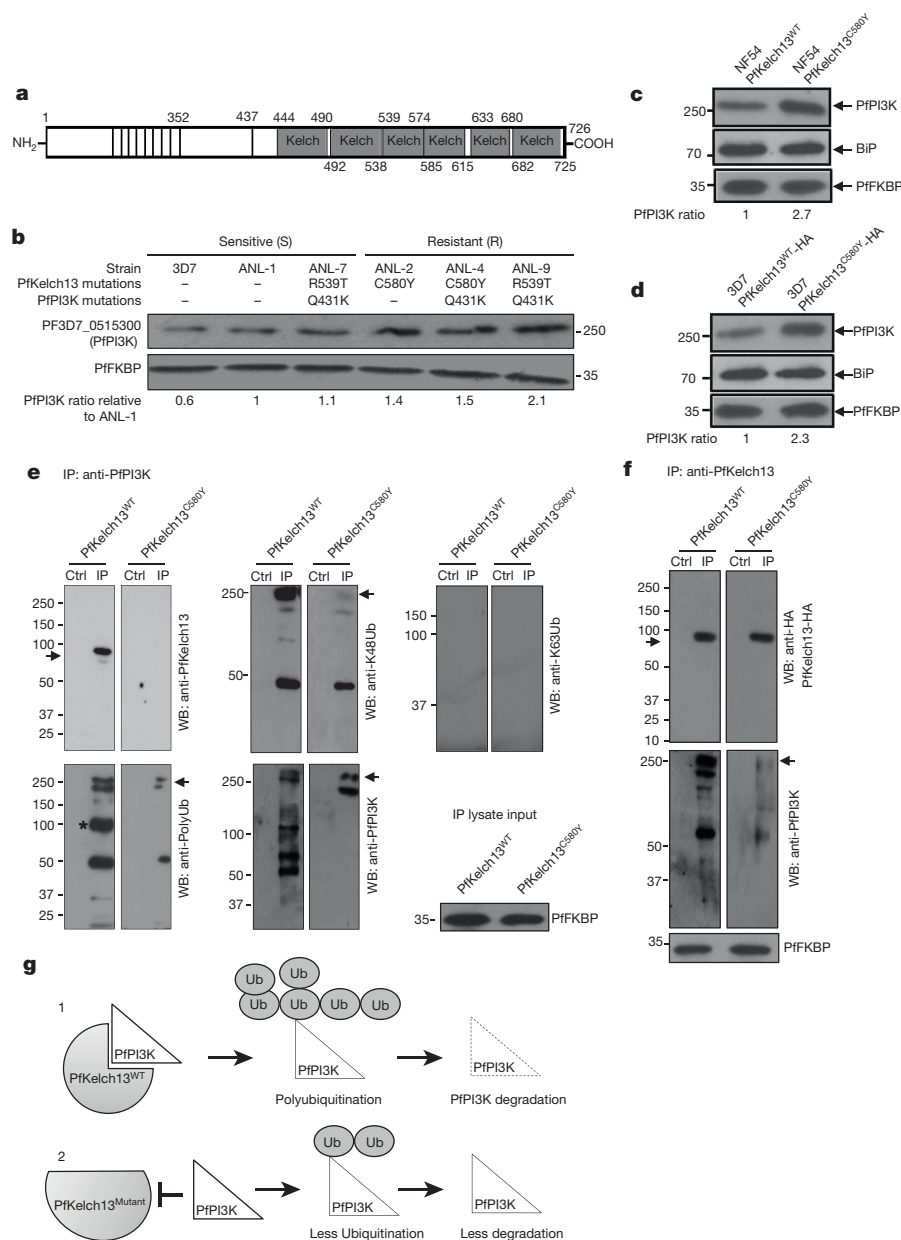
(as measured by ring-stage survival assays (RSA)[22]; Extended Data Fig. 5a) in both clinical and genetically engineered laboratory strains.

To elevate PI3P in the absence of PfKelch13, we transgenically expressed human VPS34 (Fig. 3b and Extended Data Fig. 5b–d). Human VPS34 synthesizes only PI3P and in the absence of PfKelch13 mutation increased ring PI3P and resistance in two independent 3D7 transgenic lines (Pf$^{VPS34-myc1}$ and Pf$^{VPS34-myc2}$); but a catalytically inactive VPS34 or a control reporter had no effect (Fig. 3c, d and Extended Data Fig. 5b–e). Introduction of PfKelch13^{WT} into VPS34 transgenic parasites reduced PI3P and RSA while the PfKelch13^{C580Y} mutation reciprocally increased cellular PI3P and RSA levels (Fig. 3c, d and Extended Data Fig. 5b–f). The best fit line in Fig. 3d is closely comparable to that in Fig. 3a, suggesting the VPS34 transgenic lines remained responsive to PfKelch13. These data indicate that elevated PI3P levels confer artemisinin resistance (Fig. 3e). Notably, a ~2.5-fold increase in PI3P induced greater than one order of magnitude change in resistance suggesting that as PI3P is a signalling lipid, small changes in its levels may induce downstream activation of large amplitudes of resistance (Fig. 3e).

We next queried whether additional cellular components could also influence PI3P and RSA levels. PI3K–AKT is a primary signal transduction pathway in eukaryotes but its role is poorly understood in malaria parasites. In addition to PI3P, PfPI3K can also synthesize PI3,4,5P3 (ref. 12), which in most cells is needed for activation of AKT[23]. *P. falciparum* has an orthologue of AKT (PfAKT/PF3D7_1246900; Extended Data Fig. 6a). However PfAKT appears different from its mammalian counterparts because it lacks a PH domain and a conserved Ser473. Rather unexpectedly, we found that DHA blocks cellular PfAKT activity (Fig. 4a) but did not inhibit purified PfAKT (Fig. 4b). As it targets PfPI3K (Fig. 1), we reasoned that in parasites DHA may block PfAKT through inhibition of PfPI3K. Although PfAKT lacks a PI3,4,5P3-binding PH domain, it may function through a calmodulin-binding PH domain protein[24] as PfAKT contains a calcium/calmodulin activator domain. As indicated earlier, low levels of PI3,4,5P3 are made by PfPI3K[12] (and in this regard, PfPI3K is also different from its closest homologue VPS34 which produces solely PI3P). Transgenic elevation of PfAKT (Extended Data Fig. 6b) induced a ~1.8-fold elevation of PI3P (presumably stimulated by feedback mechanisms)

Figure 3 | Elevated PI3P and artemisinin-resistance in presence and absence of PfKelch13 mutations. **a**, RSA versus PI3P levels in clinical (triangle) and laboratory (circle) strains. Numbers 1–4 on the graph indicate sensitive strains ANL-1, 3D7, NF54-PfKelch13[WT], 3D7-PfKelch13[WT]-HA with RSA ≤ 1. ANL-7 contaminated with (R539T) resistant parasites (asterisk) show intermediate PI3P/RSA levels. 3D7-PfKelch13[C580Y]-HA contains a chromosomal wild-type copy of PfKelch13 and thus shows lower PI3P and RSA (49 and 9) than NF54-PfKelch13[C580Y] (59 and ~13). **b**, 3D7 expressing transgenic myc-tagged VPS34 (left) or VPS34-catalytically dead mutant (right); loading control PfFKBP; raw data in Supplementary Data 2. Molecular weights, in kDa. Data are representative of three experimental replicates. **c**, PI3P levels in indicated strains. **d**, PI3P, RSA in VPS34 strains respond to PfKelch13. 1–4: sensitive strains ANL-1, Pf[VPS34-MUTANT-myc], SS-EEA1[R1374A]-mCherry and 3D7. In **a**, **c**, **d**, mean from three experimental replicates (each in triplicate); **a**, RSA, s.d. <0.5; **c**, PI3P, s.d. <8 (shown by error bars). **e**, A model for PI3P-induced artemisinin resistance. Further details in Extended Data Fig. 5.

e Dysregulation of proteostatic control of PfPI3K by PfKelch13 mutation elevates PI3P to induce artemisinin resistance

Figure 4 | PI3P-mediated resistance linked to GWAS. a, b, Effect of DHA on cellular (**a**) and immunopurified (**b**) PfAKT. **c**, Expression of PfAKT-GFP in *Pf* 3D7. Loading controls, PfFKBP, erythrocyte band 3. Molecular weights in kDa. RSA, PI3P, as indicated. In **a–c**, mean of three experimental replicates (each with triplicate data points). AKT activity and RSA s.d. <0.5; PI3P s.d. <5. In **b** and **c**, western blots are representative of three experimental replicates; raw data in Supplementary Data 2. **d**, PI3P-mediated resistance pathways. Bold, lipid (PI3P) or protein components from this study and GWAS; non-bold indicates GWAS only[6,7]. Solid lines, established relationships; broken/dotted lines, partial functional validation; broad grey arrows, *in silico* prediction. Additional details in Extended Data Fig. 6.

and RSA of 6.5, which is comparable to the resistance level seen in one of the clinically derived lines (ANL-4; Fig. 4c).

Evidence that PfAKT is a major downstream effector of PfPI3K is still lacking. Nonetheless, the data suggest that secondary genes may influence levels of PfPI3K/PI3P product perhaps explaining why AKT (and other kinases) are also under artemisinin-selective pressure[6]. Differences between the pools of available substrates (phosphatidylinositol; PI, ATP) extent of ubiquitination (and additional unknown regulation) of PfPI3K may also be influenced by complex lipid, biosynthetic, transport and signalling networks (Fig. 4d) that vary across genetic backgrounds. Therefore absolute levels of PfPI3K across non-isogenic strains may not provide an accurate measure of PfPI3K activity/PI3P product and artemisinin resistance. Genes involved in lipid uptake[25] are also implicated in resistance[6] and serum lipids are known to be essential for intracellular blood stage parasite growth[26]. This may explain the wide range of RSA values displayed by resistant, clinical strains and why two resistant strains like ANL-2 and ANL-4 bear the same mutation C580Y and equivalent levels of PfPI3K, but different amounts of PI3P product and thus different levels of resistance.

DHA has many other targets in the later trophozoite parasite stage[27–29], but PfPI3K is the only known target in early ring stages: clinical resistance to artemisinins develops in early rings (and in absence of haemoglobin digestion associated with either late rings or trophozoites). The two most prevalent mutations (C580Y and R539T) associate with elevation of PI3P. All resistance mutations of PfKelch13 are limited to the Kelch propeller domain, suggesting they too regulate PfPI3K activity to contribute to differences in RSA. The molecular identities of other putative PfKelch13 targets and the ubiquitin ligase activity of the complex have yet to be investigated. PI3P may influence host remodelling, and functions of the apicoplast and food vacuole[11–13] as well as cell survival through redox, transcriptional and DNA repair pathways (grey arrows, Fig. 4d), all of which have also been implicated in artemisinin resistance[5,6,9,30]. Sustained, selective targeting of the PfPI3K and/or its regulation will be critical to developing new therapies that allow eliminating artemisinin resistance and rendering effective malaria control.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. The World Health Organization. World Malaria Report 2012. http://www.who.int/malaria/publications/world_malaria_report_2012/en/.
2. Noedl, H. et al. Evidence of artemisinin-resistant malaria in western Cambodia. N. Engl. J. Med. **359,** 2619–2620 (2008).
3. Dondorp, A. M. et al. Artemisinin resistance in Plasmodium falciparum malaria. N. Engl. J. Med. **361,** 455–467 (2009).
4. Dondorp, A. M. et al. The threat of artemisinin-resistant malaria. N. Engl. J. Med. **365,** 1073–1075 (2011).
5. Cheeseman, I. H. et al. A major genome region underlying artemisinin resistance in malaria. Science **336,** 79–82 (2012).
6. Miotto, O. et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nature Genet. **45,** 648–655 (2013).
7. Ariey, F. et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature **505,** 50–55 (2014).
8. Ashley, E. A. et al. Spread of artemisinin resistance in Plasmodium falciparum malaria. N. Engl. J. Med. **371,** 411–423 (2014).
9. Takala-Harrison, S. et al. Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in Southeast Asia. Proc. Natl Acad. Sci. USA **110,** 240–245 (2013).
10. Takala-Harrison, S. et al. Independent emergence of Plasmodium falciparum artemisinin resistance mutations in Southeast Asia. J. Infect. Dis. **211,** 670–679 (2015).
11. Bhattacharjee, S., Stahelin, R. V., Speicher, K. D., Speicher, D. W. & Haldar, K. Endoplasmic reticulum PI3P lipid binding targets malaria proteins to the host cell. Cell **148,** 201–212 (2012).
12. Tawk, L. et al. Phosphatidylinositol 3-phosphate, an essential lipid in Plasmodium, localizes to the food vacuole membrane and the apicoplast. Eukaryot. Cell **9,** 1519–1530 (2010).
13. Vaid, A., Ranjan, R., Smythe, W. A., Hoppe, H. C. & Sharma, P. PfPI3K, a phosphatidylinositol-3 kinase from Plasmodium falciparum, is exported to the host erythrocyte and is involved in hemoglobin trafficking. Blood **115,** 2500–2507 (2010).
14. Shen, Y. L. J. et al. for antimicrobial hit discovery targeting metabolic networks. Proc. Natl Acad. Sci. USA **107,** 1082–1087 (2010).
15. Efferth, T. et al. Antiviral activity of artesunate towards wild-type, recombinant, and ganciclovir-resistant human cytomegaloviruses. J Mol Med. **80,** 233–242 (2002).
16. Xu, H. et al. Anti-malarial agent artesunate inhibits TNF-α-induced production of proinflammatory cytokines via inhibition of NF-κB and PI3 kinase/Akt signal pathway in human rheumatoid arthritis fibroblast-like synoviocytes. Rheumatol. **46,** 920–926 (2007).
17. Cheng, C. et al. Anti-malarial drug artesunate attenuates experimental allergic asthma via inhibition of the phosphoinositide 3-kinase/Akt pathway. PLoS ONE **6,** e20932 (2011).
18. Zhang, D. D., Lo. S. C., Cross J. V., Templeton D. J. & Hannink M.. Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex. Mol. Cell. Biol. **24,** 10941–10953 (2004).
19. Chotivanich, K. et al. Laboratory detection of artemisinin-resistant Plasmodium falciparum. Antimicrob. Agents Chemother. **58,** 3157–3161 (2014).
20. Ghorbal, M. et al. Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nature Biotechnol. **32,** 819–821 (2014).
21. Ikeda, F. & Dikic I.. Atypical ubiquitin chains: new molecular signals. 'Protein modifications: beyond the usual suspects' review series. EMBO Rep. **9,** 536–542 (2008).
22. Witkowski, B. et al. Novel phenotypic assays for the detection of artemisinin-resistant Plasmodium falciparum malaria in Cambodia: in-vitro and ex-vivo drug-response studies. Lancet Infect. Dis. **13,** 1043–1049 (2013).
23. Manning, B. D. & Cantley, L. C. AKT/PKB signaling: navigating downstream. Cell **129,** 1261–1274 (2007).
24. Yano, S., Tokumitsu, H. & Soderling, T. R. Calcium promotes cell survival through CaM-K kinase activation of the protein-kinase-B pathway. Nature **396,** 584–587 (1998).
25. van Ooij, C. et al. Identification of a Plasmodium falciparum phospholipid transfer protein. J. Biol. Chem. **288,** 31971–31983 (2013).
26. Mi-Ichi, F., Kano, S. & Mitamura, T. Oleic acid is indispensable for intraerythrocytic proliferation of Plasmodium falciparum. Parasitology **134,** 1671–1677 (2007).
27. Eckstein-Ludwig, U. et al. Artemisinins target the SERCA of Plasmodium falciparum. Nature **424,** 957–961 (2003).
28. Klonis, N. et al. Artemisinin activity against Plasmodium falciparum requires hemoglobin uptake and digestion. Proc. Natl Acad. Sci. USA **108,** 11405–11410 (2011).
29. Cheng, Q., Kyle, D. E. & Gatton, M. L. Artemisinin resistance in Plasmodium falciparum: a process linked to dormancy? Int. J. Parasitol., Drugs and Drug Resist. **2,** 249–255 (2012).
30. Painter, H. J., Campbell, T. L. & Llinas, M. The Apicomplexan AP2 family: integral factors regulating Plasmodium development. Mol. Biochem. Parasitol. **176,** 1–7 (2011).

METHODS

Cloning and generation of constructs for *P. falciparum* transfection. All constructs used for generating transgenic *P. falciparum* 3D7 parasites were assembled in different plasmids and finally cloned into pA150 or pA156[11]. This resulted in the expression of the proteins under the calmodulin (*cam*) promoter. The generation of constructs pA156 (SS-EEA1[WT]-mCherry) and pA156 (SS-EEA1[WT]-mCherry) have been described earlier[11].

For the generation of constructs pA150 (PfKelch13[WT]-HA), *pfkelch13* (PF3D7_1343700) was amplified from *P. falciparum* genomic DNA using PfKelch13AvrIIF and PfKelch13HAXhoIR; digested with AvrII/XhoI and cloned into corresponding sites of pA150. The construct pA150 (PfKelch13[C580Y]-HA) was generated using overlapping PCR strategy. Briefly, the products of PCR1 (using PfKelch13AvrIIF and C580Yreverse) and PCR2 (using C580Yforward and PfKelch13HAXhoIR) reactions were used as template for the overlapping PCR3 to generate full length *pfkelch13* (with residues conferring C580Y mutation in the protein sequence), which was subsequently digested with AvrII-XhoI and cloned into pA150 to generate pA150 (PfKelch13[C580Y]-HA).

The plasmid containing human *VPS34* (Cat# SC118487) was purchased from Origene Inc. (Rockville, MD, USA). This was used as template to PCR amplify human *VPS34* using HsVPS34AvrIIF and HsVPS34mycXhoIR that was subsequently digested and cloned to the AvrII/XhoI site of either pA156 (with *bsd* resistance cassette) or pA150 (with *hdhfr* resistance cassette) to generate pA156 (VPS34-myc1) and pA150 (VPS34-myc2), respectively. The construct pA150 (VPS34[MUTANT-myc1]) was generated using overlapping PCR strategy. Briefly, the products of PCR1 (using HsVPS34AvrIIF and VPS34-742AAA745R) and PCR2 (using VPS34-742AAA745F and HsVps34mycXhoIR) reactions were used as template for the overlapping PCR3 to generate full length *VPS34[mutant-myc1]* (with residues [742]DRH[745] changed to AAA in the protein sequence), which was subsequently digested with AvrII/XhoI and cloned into pA150 to generate pA150 (VPS34[MUTANT-myc1]).

The construct pA150 (PfAKT-GFP) was generated as follows. Briefly, the full-length *pfakt* was amplified from *P. falciparum* genomic DNA using AKTAvrII_F and AKTBglII_R primers and cloned into AvrII/BglII site of pA150 (SSGFP)[11]. This resulted in an in-frame fusion with *gfp* at the 5′ end.

A single amino acid mutation replacing cysteine at position 580 with tyrosine (C580Y) in PfKelch13 (PF3D7_1343700) was generated using CRISPR–Cas9 technology as described elsewhere[20].

Sequence analysis of *pfkelch13* and *pfpi3k* polymorphisms in the clinical strains. Clinical strains were first propagated in tissue culture and genomic DNA isolated using Quick-gDNA Blood MiniPrep kit from ZYMO Research following manufacturer's instructions.

For the sequence analysis for polymorphism in *pfkelch13*, either the full length *pfkelch13*, the full-length gene was amplified using Kelch13_F and Kelch13_R primers, or specific regions were amplified using Kelch13-1, Kelch13-2, Kelch13-3, Kelch13-4, Kelch13-5 and Kelch13-6 primers and analysed.

To check for *pfpi3k* polymorphism in clinical strains, specific regions of *pfpi3k* were amplified using PfPI3K-I682T-F/ PfPI3K-I682T-R, PfPI3K-Q431K-F/ PfPI3K-Q431K-R, PfPI3K-Y1330C-F/ PfPI3K-Y1330C-F primer pairs and sequenced.

Generation of anti-peptide antibodies. All anti-peptide antibodies were custom-generated by Thermo Scientific (Rockford, IL, USA). Antibodies against *P. falciparum* PI3K (PfPI3K, PF3D7_0515300) were generated in guinea pigs and rabbits against the C-terminal peptides CVDKLHEWALNWK and CVLKVQEKFRLD-LNDE, respectively. Antibodies were also custom-generated in rat, against the peptide RKRFDEERLRFLQEIDKI and corresponding to the amino acids 299–316 of PfKelch13 (PF3D7_1343700). Antibodies to BiP (PF3D7_0917900) were generated against the peptide DYFIKMFKKKNNIDLRTDKR corresponding to amino acids 261–280. For the generation of rabbit antibodies to PfAKT (RAC-beta serine/threonine protein kinase, PF3D7_1246900), amino acids corresponding to 229–246 (KKDKKIINLKKYNNAHRD) were selected. Antibodies were also custom-generated in rats, against the peptide REIVGDNTIEKKTEKALRE corresponding to amino acids 90–108 of PfExp-2 (PF3D7_1471100).

All guinea pig antibody generations involved collection of control sera on day 0 followed by immunization on day 1 with 0.1 mg of keyhole limpet haemocyanin (KLH)-conjugated peptide with complete Freund's adjuvant (CFA) subcutaneously at 4 sites. Booster immunizations were performed on days 21 and 42 with 50 μg of KLH-peptide with incomplete Freund's adjuvant subcutaneously. Test bleeds were checked by western blotting with uninfected and infected erythrocytes and corresponding animals were finally boosted again on day 62 followed by exsanguination bleed and peptide-affinity purification of the antibodies.

For antibody generation in rabbits, control sera were collected on day 0 and animals were immunized on day 1 with 500 μg of KLH-conjugated peptides subcutaneously at 10 sites. Animals were again boosted on days 14 and 28 with 250 μg

antigen and sera collected on day 35 to check the reactivity. Finally, after a booster immunization of selected animals on day 56/58, production bleed was collected on day 72 and processed for peptide-affinity purification of the antibodies.

The reactivity of all custom-generated antibodies were confirmed by western blots using uninfected and infected red cells and used either for immunofluorescence, immunoprecipitation assays or western blot analyses.

Parasite culture, transfection and live cell imaging. *P. falciparum* 3D7 parasites were propagated in A+ human erythrocytes (bought from Biochemed Services, Winchester VA, USA: protocol approved by the University of Notre Dame) in RPMI supplemented with 0.5% Albumax II, 0.2 mM hypoxanthine, 11 mM Glucose, 0.17% NaHCO$_3$, 10 μg/ml gentamycin and maintained at 37 °C in 5% C0$_2$ in a humidified incubator. Synchronous *P. falciparum* 3D7 parasites in culture were transfected with indicated plasmids by standard procedures as described earlier[11]. Forty-eight hours after transfection, the cultures were selected either with 2.5 nM WR99210 (Jacobus Pharmaceuticals; for transfections involving pA150) or 1 μg ml^{-1} blasticidin hydrochloride until stable cells lines were obtained. Culture-adapted, clinical field strains of *P. falciparum* as well as NF54 parasites expressing PfKelch13[WT] and PfKelch13[C580Y] (ref. 20) were propagated in the same media at 37 °C in 90% N$_2$, 5% 0$_2$ and 5% C0$_2$ in a humidified incubator. Cultures were monitored daily by Giemsa staining of methanol-fixed smears and fed as necessary. Parasitemia were usually maintained below 12% for healthy culture growth.

Patient sample collection and short term cultures. After hospital admission and informed consent, 10 ml of whole *P. falciparum*-infected blood was drawn from adults at multiple field sites in western Cambodia. White blood cells were removed over a CF11 column. The infected red blood cells were subjected to short term *in vitro* culture in RPMI1640 containing 20% human serum or Albumax II (0.5%) for a maximum of eight weeks. This study was approved by the Oxford Tropical-medicine Research Ethical Committee (OXTREC) as well as the Ministry of Health in Cambodia; the trial was registered under NCT00493363. The study was also approved by the University of Notre Dame.

Treatments with PI3K inhibitors, wortmannin, LY294002, the inactive orthologue LY303511 and artemisinins. Highly synchronized parasites at schizont stages from *P. falciparum* 3D7 (transgenically expressing secretory SS-EEA1[WT]-mCherry) were purified using percoll density gradients. Purified schizonts were then allowed to invade fresh batch of red blood cells for 6 h. The resulting intracellular rings were exposed to mock treatment (0.1% DMSO) or the following compounds: wortmannin, LY294002, inactive LY294002 analogue LY303511 (from Selleckchem), artemisinin, artesunate or dihydroartemisinin (DHA) (from Sigma–Aldrich) at the indicated concentrations in DMS0 (0.1%). Drug treatments were carried out as indicated for 30 min or 4 h. Cells were subsequently processed for live cell imaging, immunofluorescence assay (IFA) and western blotting. Drugs were removed by washing thrice with serum-free RPMI, and infected erythrocytes were imaged after 1 h. In addition, where indicated parasite growth in culture was monitored 24 h later in Giemsa smears.

Imaging of live and cells by fluorescence microscopy and quantitative analyses. For live cell imaging, parasites were processed as described previously[31]. In both live and fixed cells, the parasite nuclei were stained with Hoechst 33342 and cells were imaged with a ×100, NA 1.4 objective on an Olympus IX inverted fluorescence microscope with a temperature controlled stage and a Photometrix cooled CCD camera (CH350/LCCD) driven by DeltaVision software from Applied Precision (Seattle, WA).

Quantitative analysis for the fraction of perinuclear SS-EEA1[WT]-mCherry fluorescence was undertaken using DeltaVision software. Briefly, consecutive 0.2 micron optical z-sections per cell were observed for either perinuclear/ER-like fluorescence pattern or peripheral/PV-like pattern under control or drug-treated condition and represented graphically. 400 optical sections were quantified per treatment (10 in-focus optical sections of the parasite per infected cell). While perinuclear/ER fluorescence was an indication of no observable effect on the PI3P level, peripheral/PV fluorescence was an indication of a reduction in PI3P level on drug treatment.

Immunoprecipitation of PfPI3K and PI3-kinase assay. PfPI3K (PF3D7_0515300) was immunoprecipitated from parasite stocks previously frozen at −80 °C. Proteins were extracted 1 h at 4 °C using extraction/lysis buffer containing 10 mM Tris. HCl, pH 7.5, 100 mM NaCl, 5 mM EDTA, 1% Triton X-100, 100 μM sodium orthovanadate, 20 μM sodium fluoride, 20 μM β-glycerophosphate, and 1× protease inhibitor cocktail, Roche Diagnostics; 10% glycerol). After removal of insoluble debris, the extract was incubated with guinea pig anti-PfPI3K for 12 h at 4 °C. Protein A agarose beads (pre-equilibrated in lysis buffer) was then added and additionally incubated for 6 h. Immune complexes bound to beads were extensively washed with extraction buffer. PfPI3K-bound beads were either processed for western blotting or kinase activity assay. For western blotting beads were directly boiled in reducing SDS–PAGE sample buffer, resolved by SDS–PAGE and probed with rabbit anti-PfPI3K. Kinase assay of PfPI3K in

immunoprecipitated beads was measured using Class III PI3-Kinase kit from Echelon Biosciences (K-3000) following the manufacturer's instructions. Mean (s.d.) from three experimental replicates, each containing duplicate data points are shown.

Mammalian kinase assays. The kinase activity of mammalian kinases was performed by Life Technologies (Grand Island, NY, USA) in the absence or presence of 10 μM DHA, through SelectScreen Biochemical Kinase Profiling Service, and expressed as percentage inhibition.

Immunolocalization by indirect immunofluorescence assays (IFA). Indirect immunofluorescence assays (IFAs) were performed on non-transfected *P. falciparum* 3D7 or transgenic parasites fixed with glutaraldehyde/paraformaldehyde as described previously[32] using the following antibody concentrations- rabbit anti-PfPI3K (custom-made, 20 μg ml^{-1}), guinea pig anti-Exp-2 (custom-made, 20 μg ml^{-1}) and mouse anti-myc antibodies from Abcam (Cambridge, MA, USA). The appropriate FITC- or TRITC labelled secondary IgG antibodies (ICN Biochemicals) were used at 1:200 dilution. Parasite nuclei were stained with 5 mg ml^{-1} Hoechst 33342 (Molecular Probes) and slides were mounted with DABCO.

Immunoprecipitation and western blotting. PfPI3K was immunopurified from *P. falciparum* 3D7 transgenic parasites (3D7-PfKelch13^{WT}-HA and 3D7-PfKelch13^{C580Y}-HA) using guinea pig anti-PfPI3K as described above. Immunoprecipitations using Protein A agarose beads were used as negative control. Western blots were subsequently performed using custom-generated rabbit anti-PfPI3K, rat anti-PfKelch13 antibodies, commercial rabbit anti-polyubiquitin (MBL-PW8810), K48-linkage specific polyubiquitin (Cell Signaling Technology-4289S), K63-linkage specific polyubiquitin (Cell Signaling Technology-5621S) or mouse monoclonal anti-HA tag antibody from (Abcam-ab18181) antibodies.

To measure the relative amount of PfPI3K expression in *P. falciparum* 3D7 and clinical isolates, 4×10^6 infected erythrocytes permeabilized with 0.01% saponin followed by extensive washing with PBS to remove haemoglobin. Cells were solubilized in Laemmli's sample buffer, resolved by SDS–PAGE and western blotting performed using guinea-pig anti-PfPI3K, followed by HRP-conjugated donkey-anti guinea pig antibodies from Jackson ImmunoResearch (West Grove, PA, USA). Blots were developed using chemiluminescence assay kit from Thermo Scientific (Rockford, IL, USA). Antibodies to the parasite protein PfFKBP served as a loading control. The intensities of PfPI3K and PfFKBP signals were quantitated by densitometry and expressed as a ratio.

The relative amount of PfPI3K expression in the transgenic parasites NF54-PfKelch13^{WT}, NF54-PfKelch13^{C580Y}, 3D7-PfKelch13^{WT}-HA and 3D7-PfKelch13^{C580Y}-HA was measured by resolving the saponin-permeabilized pellet from 2×10^7 parasites by SDS–PAGE and western blotting using guinea pig anti-PfPI3K followed by HRP-conjugated donkey-anti guinea pig antibodies and chemiluminescence detection. Antibodies to parasite proteins BiP and PfFKBP were used as loading controls. Antibodies to BiP were custom generated, while antibodies to PfFKBP were kindly provided by Nirbhay Kumar. The intensities of PfPI3K and PfFKBP signals were quantitated by densitometry and expressed as a ratio.

Transgenic *P. falciparum* 3D7 parasites, either expressing myc-tagged VPS34 (VPS34-myc1 and VPS34-myc2), those expressing PfKelch13^{WT}-HA/PfKelch13^{C580Y}-HA, as well as those expressing PfKelch13^{WT}-HA/PfKelch13^{C580Y}-HA in VPS34-myc1 background were also checked for the expression by western blotting. Primary antibody either involved mouse anti-myc antibodies from Abcam (Cambridge, MA, USA) or rabbit anti-HA antibodies (Thermo Fisher Scientific, Grand Island, NY, USA), followed by HRP-conjugated secondary antibody from Bio-Rad (Hercules, CA, USA). Blots were developed using chemiluminescence detection kit. Antibodies to parasite PfFKBP or BiP were used as loading controls.

Expression of SS-EEA1^{R1374A}-mCherry was detected using Living Colours polyclonal rabbit anti-DsRed2 (Clontech, Mountain View, CA, USA).

Expression of PfAKT-GFP in transgenic parasites was confirmed by western blotting of uninfected erythrocytes, non-transfected 3D7 parasites and transgenic (PfAKT-GFP) parasites, using rabbit anti-GFP antibodies from Thermo Fisher Scientific (Grand Island, NY, USA) and HRP-conjugated secondary goat anti-rabbit antibodies from Bio-Rad (Hercules, CA, USA). Antibodies to parasite PfFKBP and human band 3 were used as loading controls.

Model building, docking and molecular dynamics (MD) simulations. The homology model of *P. falciparum* PfPI3K was build based on the structure of the class III PI3-kinase from *Drosophila* (PDB code 2X6F)[33], using the multiple threading alignment in I-TASSER[34,35] using additional information for the definition of the active site from the structure of the human class III PI3K (PDB code 3IHY).

The known PI3K inhibitor, LY294002 (ref. 36) was docked to the model using GlideXP[37] to generate the starting structure for model refinement. The initial structure for the PI3K-dihydroartemisinin (DHA) complex was generated by

removing the LY294002 ligand from the refined structure and docking DHA to this model using GlideXP.

This model of PfPI3K then served as the initial structure for 20 ns MD simulations. The model in these simulation consisted of the protein, surrounded by a periodic box of TIP3P[38] water molecules that extended 10 Å from the protein. Na$^+$ counter ions were placed by the LEaP module of AMBER12[39] to neutralize the system. Ionizable residues were set to their normal ionization states at pH 7.

The MD simulations were carried out using the PMEMD module of AMBER12 (ref. 39). The protein was modelled using the ff03.r1 version of the AMBER force field while the DHA ligand was represented by the GAFF force field[40,41]. Atom-centred partial charges were generated based on B3LYP/6-31G* optimized geometries using RESP[42,43] as implemented in the antechamber module of AMBER12. A time step of 2 fs, combined with the SHAKE algorithm[44] to constrain all bonds involving hydrogen atoms, was chosen. A non-bonded cutoff of 10.0 Å was used, and the non-bonded pair list was updated every 25 time steps.

The temperature (300 K) and pressure (1 atm) of the NPT ensemble were controlled by Langevin dynamics and isotropic position scaling[45] respectively. Long-range interactions were treated by the Particle-Mesh-Ewald (PME)[46,47] methods with a grid spacing of ~1 Å and a fourth-order B-spline interpolation to compute the potential and forces in between grid points. The trajectories were analysed using the PTRAJ module of AMBER12. The model of the PfPI3K-DHA complex was simulated analogously for a total time of 20 ns.

The same MD protocol was used for the study of the complexes of artelinic acid, artemisone, and deoxyartemisinin in the PfPI3K model as well as DHA in VPS34 and p110γ. In short, initial structures were generated by docking of the small molecules to the PfPI3K model, human VPS34 (PDB code 3IHY), and human p110γ (PDB code 4ANV). After manual inspection of the generated poses, the initial models were subjected to 20 ns MD simulations as described above.

Assay for measuring cellular and immunoprecipitated PfAKT activity. *P. falciparum* 3D7 early-ring stage parasites were synchronized by sorbitol treatment and exposed either to mock treatment: 0.1% DMSO (control/mock) or 4 nM DHA for 3 h. An aliquot was washed using serum-free RPMI and returned to culture for 6 h (washout). AKT activity was measured in mock-treated, 4 nM DHA-treated and post-washed DHA treated cells (4×10^7) using the K-LISA Akt Activity Kit (CBA019) from Calbiochem (Darmstadt, Germany) following manufacturer's instructions. Briefly, phosphorylation of AKT biotinylated peptide substrate was detected with phosphoserine antibodies. Serial dilutions of active human Akt1/PKBα (Millipore) were used as standard and absorbance read at 450 nm on a plate reader. Mean (s.d.) from three experimental replicates, (each with triplicate data points) are shown.

Endogenous PfAKT was also immunopurified from *P. falciparum* 3D7 following saponin treatment (as described earlier) and the resulting parasite pellet extracted with extraction buffer (25 mM Tris.HCl pH 8.0, 150 mM NaCl, 1% (v/v) Triton X-100, 1% (w/v) deoxycholate-Na, 2 mM EDTA containing complete protease inhibitor for 1 h at 4 °C. PfAKT was immunopurified using 10 μg of custom-generated antibodies (for 2 h at 4 °C, under shaking conditions) followed by protein A agarose beads for 1 h. Protein A agarose beads incubated with parasite lysate in absence of antibodies served as negative control. Immunopurified PfAKT-beads and control beads were extensively washed in PBS, pH 7.4 and processed further for the measurement of activity in the absence or presence of 4 nM DHA as described previously. Immunoprecipitation was also confirmed by western blotting using anti-PfAKT antibodies.

Measurement of PI3P level in clinical and laboratory strains. Early ring stage parasites from *P. falciparum* 3D7 laboratory strains (non-transfected or transgenic) and clinical strains were tightly synchronized with sorbitol treatment and lysed with saponin to remove haemoglobin. Resulting pellets (corresponding to 6×10^7 parasites) were washed with PBS. Lipids were extracted and PI3P level assessed (in triplicates) using Echelon PI3P Mass ELISA Kit (K-3300) following the manufacturer's instructions.

PI3P level was also measured in 3D7 parasites under control condition, after 3 h treatment with 4 nM DHA, as well as after 6 h in culture following DHA washout. Mean (s.d.) from three experimental replicates, (each with triplicate data points) are shown.

Ring-stage survival assay (RSA). *In vitro* RSA was assessed as described by elsewhere[22]. Briefly, tightly synchronized early ring parasites (0–3 h post-invasion) were exposed to 700 nM DHA or 0.1% DMSO (control) for 6 h, washed thrice with serum-free RPMI and returned to culture for 66 h. Blood smears were prepared and stained with Giemsa. Each sample was done in duplicate and 10,000 erythrocytes were assessed independently by light microscope. Parasite survival rates (%RSA) were expressed as a percentage by comparing the number of viable parasites between the drug-treated and untreated control. Mean (s.d.) of RSA based on four replicated carried out by two independent laboratory personnel.

Sample size. No statistical methods were used to predetermine sample size.

31. Osborne, A. R. *et al.* The host targeting motif in exported *Plasmodium* proteins is cleaved in the parasite endoplasmic reticulum. *Mol. Biochem. Parasitol.* **171,** 25–31 (2010).

32. Tonkin, C. J. *et al.* Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol. Biochem. Parasitol.* **137,** 13–21 (2004).

33. Miller, S. *et al.* Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34. *Science* **327,** 1638–1642 (2010).

34. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9,** 40 (2008).

35. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5,** 725–738 (2010).

36. Vlahos, C. J., Matter, W. F., Hui, K. Y. & Brown, R. F. A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). *J. Biol. Chem.* **269,** 5241–5248 (1994).

37. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **49,** 6177–6196 (2006).

38. Jorgensen, W. L., Chandrasekhar, J., Buckner, J. K. & Madura, J. D. Computer simulations of organic reactions in solution. *Ann. NY Acad. Sci.* **482,** 198–209 (1986).

39. Case, D. A. *et al.* AMBER 12. (http://www.ambermd.org) (Univ. of California, San Francisco, 2012).

40. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25,** 247–260 (2006).

41. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25,** 1157–1174 (2004).

42. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97,** 10269–10280 (1993).

43. Fox, T. & Kollman, P. A. The application of different solvation and electrostatic models in molecular dynamics simulations of ubiquitin: how well is the X-ray structure "maintained"? *Proteins* **25,** 315–334 (1996).

44. Ryckaert, J. P., Ciccotti, G. & Berendsen, J. C. H. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

45. Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **65,** 1409–1419 (1988).

46. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

47. Petersen, H. G. Accuracy and efficiency of the particle-mesh-Ewald method. *J. Chem. Phys.* **103,** 3668–3679 (1995).

a

b DAPI + Brightfield | SS-EEA1^WT-mCherry | Merge

c

d DAPI + Brightfield | SS-EEA1^WT-mCherry | Merge

e SS-EEA1^WT-mCherry

f PfPI3K / PFE0765w (PF3D7_0515300)

Peptide (Rabbit): CVLKVQEKFRLDLNDE
Peptide (Guinea pig): CVDKLHEWALNWK
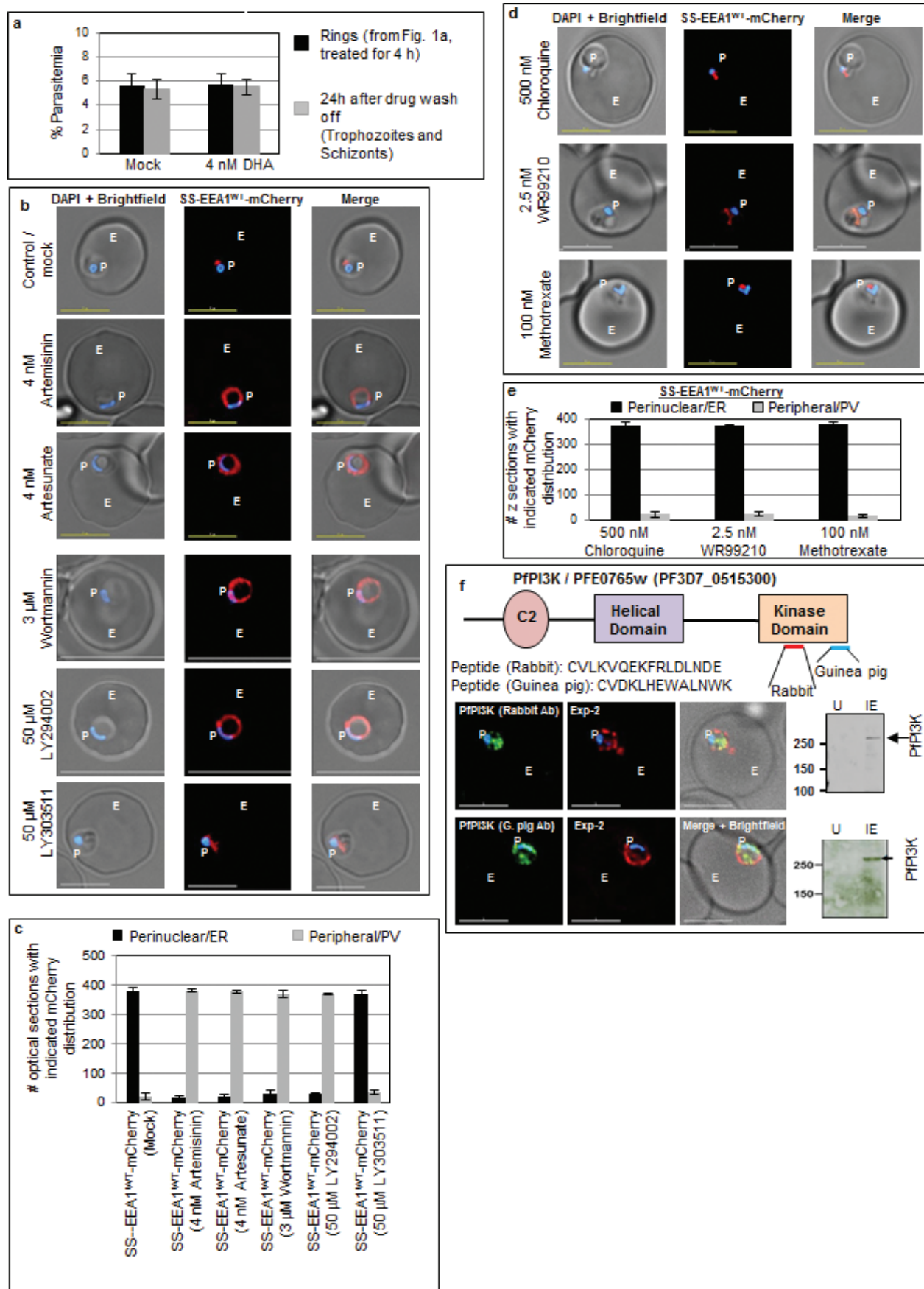
Extended Data Figure 1 | Ring Parasite PI3P and PfPI3K: effect of inhibitors, and characterization of antibodies. a, Rings at 6 h post-invasion, were either mock treated or exposed to 4 nM DHA for 4 h. An aliquot was washed in serum-free RPMI and cultured for 24 h. Parasitemia (ring and trophozoite stages) were determined by Giemsa staining and light microscopy. Mean (error bars s.d. <2) from three experimental replicates are shown. **b**, Effect of artemisinins, deoxyartemisinin, PI3K inhibitors wortmannin and LY294002 (and its inactive orthologue LY303511) on PI3P level, as observed by redistribution of the PI3P reporter in transgenic 3D7 parasites expressing SS-EEA1^{WT}-mCherry (red). Blue, parasite nucleus, P, parasite, E, erythrocyte; scale bar, 5 μm. Fluorescence and phase-merged images are as indicated. **c**, Quantitative analysis of the data in **b**. Mean (error bars s.d. <21) from three experimental replicates are shown. **d**, Effect of anti-folates and chloroquine on PI3P production, as observed by redistribution of the PI3P reporter in transgenic 3D7 parasites expressing SS-EEA1^{WT}-mCherry (red). Blue, parasite nucleus, P, parasite, E, erythrocyte; scale bar, 5 μm. Fluorescence and phase-merged images are as indicated. **e**, Quantitative analysis of the data in **d**. Mean (error bars s.d. <27) from three experimental replicates are shown. As indicated in Fig. 1, the PI3P reporter SS-EEA1^{WT}-mCherry (red) detects PI3P in punctate, perinuclear ER domains in 6 h rings[11] (top panel) or reduced PI3P when the reporter diffuses to the PV (periphery). **f**, Schematic showing structural features of PfPI3K. Custom antibodies were generated in rabbits and guinea pigs against the indicated peptides. Western blots indicate that both antibodies specifically recognize a ~255 kDa protein in infected (IE) but not uninfected (U) erythrocytes and are representative of 10 experimental replicates; raw data in Supplementary Data 2. Indirect immunofluorescence assays using rabbit (top panel) and guinea pig (bottom panel) antibodies confirm the PfPI3K (green) is localized in the parasite. Exp-2 (red) marks the parasite boundary. Blue, parasite nucleus, P, parasite, E, erythrocyte; scale bar, 5 μm. Fluorescence and phase-merged images are shown. Data are representative of 10 experimental replicates.

```
PfPI3K      DDLYFFQYKRSSDEK------LLNTDLSNDSNDMIHYIDDSK------------------ 1556
HsVPS34     HDLKVWPNVEADGSEPTKTPGRTSSTLSEDQMSRLAKLTKAHRQGHMVKVDWLDRLTFRE 200
DmVPS34     DDFRLYMTQTCDNIELERSAYTWQQRLLYEHPLRLANSTKMP---ELIRERHPTRTFLIV 220
                .*: .:      . ...:    .      .  *  :* .  **
PfPI3K      --NVKIERNRDNSFFSNFLQFN----DNLDFFLNATYSDEDNNYEILD------------ 1598
HsVPS34     IEMINESEKRSSNFMYLMVEFRCVKDDKEYGIVYYEKDGDESSPILTSFELVKPVDPQM 260
DmVPS34     VKNENDQSTFTLSVNEQDTPFSLTESTLQKMNRSQMKMNDRTSDYILKVSGRDEYLLGDY 280
               :  .    .   *       .        *          * .          **
PfPI3K      DSINFVQKQKIKKIKTPLILPIDPNIELLSFL---PEQSYVLRSS--------------- 1640
HsVPS34     SMENLVESKHHKLARSLRSGPSDHDLKPNAAT---RDQLNIIVSYPPTKQLTYEEQDLVW 317
DmVPS34     PLIQFLYIQEMLSDSAVPNVVLQSVYRLESYINHHNEQAMVTKRPLPKKRTVHLHKSISS 340
                :::  :.    :       :    . :       :*  :
PfPI3K      ----------LYPIVIACL-------------------------VRKKIKLYNENYNNLIIN 1667
HsVPS34     KFRYYLTNQEKALTKFLKCVNWDLPQEAKQALELLGKWKPMDVEDSLELLSSHYTNPTVR 377
DmVPS34     LWDMGNYFQLTLHSISNVNFDKTRALKVGVHVCLYHGDKKLCAQRSTDSPNGNFDTFLFN 400
                           .      .                    .. ..  ::  ..
PfPI3K      NHTFYK-------------------NDQNKDN--------IINNLSYDKSYHSYYNSQF 1699
HsVPS34     RYAVAR-----------------LRQADDEDLLMYLLQLVQALKYENFDDIKNGLEP 417
DmVPS34     DLVMDFDIQMRNLPRMTRLCIVIFEVTKMSRSKKSSNNKDIALKDVPYNKNPLAWVNTTI 460
                .:                        .:         :: : *::  .
PfPI3K      IKTLQNSFESTTSLNYHYNFLKCSNNNIFYKNKKIERIK------------PNTSIQKA 1746
HsVPS34     TKKDSQSSVSENVSNSGINSAEIDSSQIITSPLPSVSSP-------------PPASKTKE 464
DmVPS34     FDHKDILRTGRHTLYTWTYADDIQSVEVFHPLGTIEPNPRKEECALVDLTFLSSGTGTVR 520
                .  .    .     .  . .. :::
PfPI3K      FPSNENILN-----------------------RNQHVYYSNNQIVHNIKKMN------ 1775
HsVPS34     VPDGENLEQDLCTFLISRACK-------------NSTLANYLYWYVIVECEDQD------ 505
DmVPS34     YPSEEVVLQYAADREQVNRLQRQLAGPEKPIKELKEIMANYTGLDKIYEMVDQDRNAIWE 580
                *. . * :  :                    .. .  *   .  :
PfPI3K      -------------------------------------------KHKRD--DYMINEKVLP 1790
HsVPS34     -------------------------------------------TQQRDPKTHEMYLNVMR 522
DmVPS34     RRNDILRELPEELSILLHCVYWKERDDVADMWYLLKQWPLISIERSLELLDYAYPDPAVR 640
                                                            :. :    :      .:
PfPI3K      CVSNSCLG--------DKLMPSHDKMRSS------------------------------ 1811
HsVPS34     RFSQALLKGDKSVRVMRSLLAAQQTFVDR------------------------------ 551
DmVPS34     RFAIRCLHFLKDEDLLLYLLQLVQAIKHESYLESDLVVFLLERALRNQRIGHYFFWHLRS 700
                .:   *          *:  : :
PfPI3K      -----------------------------HDKMMPSHDK-----MMPSHDKLMSPHYT 1835
HsVPS34     -----------------------------LVHLMKAVQR-----ESGNRKKKNERLQA 575
DmVPS34     EMQTPSMQTRFGLLLEVYLKGCKHHVAPLRKQLHVLEKLKQGSLIAKKGSKEKVKTMLQD 760
                                             :::    .:      .:.*
PfPI3K      LMSSHDKP---------------VAPSGVSSLGEKKSKDEKKNRKKYNEIYQL--SIKKY 1878
HsVPS34     LLGDNEKMNLSDVELIPLPLEPQVKIRGIIPETATLFKSALMPAQLFFKTEDG--GKYPV 633
DmVPS34     FLRDQRNSAVFQN--IQNPLNPSFRCSGVTPDRCKVMDSKMRPLWVVFENADVNASDVHI 818
                ::  .:.        *:.     .        : :
PfPI3K      IYKAGDDLRQDHLVIQIIYVMDNIWKRYGLDLKMTLYRVLALSTDDGFIEFVDYAESISS 1938
HsVPS34     IFKHGDDLRQDQLILQIISIMDKLLRKENLDLKLTPYKVLATSTKHGFMQFIQSVPVAEV 693
DmVPS34     IFKNGDDLRQDMLTLQMLRVMDQLWKRDGMDFRMNIYNCISMEKSLGMIEVVRHAETIAN 878
                *:* ****** * :*:: :**:: :: .:*::. *. ::  ... *:::: .
PfPI3K      IKKN------YKGEIRQYFIDNSTCSSSPLGFDTEILQNFISSCAGYSVITYILGIGDRH 1992
HsVPS34     LDI--------EGSIQNFFRKYAPSENGPNGISAEVMDTYVKSCAGYCVITYILGVGDRH 745
DmVPS34     IQKEKGMFSATSPFKKGSLLSWLKEHNKPADKLNKAINEFTLSCAGYCVATYVLGVADRH 938
                :..    .    : : .      . *.    : ::: :  *****.*  **:**:.***
PfPI3K      LDNLMVTKDGRFFHIDFGYIFG---------EDPKPFSPPMKLCKEMIEAMGGAHSIGYE 2043
HsVPS34     LDNLLLTKTGKLFHIDFGYILG---------RDPKPLPPPMKLNKEMVEGMGGTQSEQYQ 796
DmVPS34     SDNIMVKRNGQLFHIDFGHILGHF KEKLGVRRERVPFVLTHDFVYVINKGFNDRESKEFC 998
                **::::.: *::******:*:*        .:  *:  .:  .:.. . * :
PfPI3K      QFLKKCCLAYKYLRYHSQLIISLLDAMCDAGLKDMKMSPELCVLKVQEKFRLDLNDEAAE 2103
HsVPS34     EFRKQCYTAFLHLRRYSNLILNLFSLMVDANIPDIALEPDKTVKKVQDKFRLDLSDEEAV 856
DmVPS34     HFQELCERAFLVLRKHGCLILSLFSMMISTGLPELSSEKD--LDYLRETLVLDYTEEKAR 1056
                .* : *  *:  ** :. **:.*: * .  :     .: :  :::.: ** .:* *
PfPI3K      IYFLSVINASVKTLFPVVVDKL-HEWALNWK- 2133
HsVPS34     HYMQSLIDESVHALFAAVVEQI-HKFAQYWRK 887
DmVPS34     EHFRAKFSEALANSWKTSLNWASHNFSKNNKQ 1088
                ::  : :.  ::    :. ::    *::: :
```

Extended Data Figure 2 | Structural analyses of PfPI3K. Sequence alignment of the putative catalytic domain of PfPI3K and its closest human orthologues human VPS34 and *Drosophila* VPS34. Asterisk symbol indicates identity; dots indicate homology. D1889 and Y1915 residues are highlighted in yellow.

a

DHA

Y1915

D1889

Artelinic acid

Y1915

R1368

D1889

Artemisone

Y1915

D1889

R1368

Deoxyartemisinin

Y1915

D1889

b

hVPS34...DHA

D761

Y670

D644

hVPS34 (gray)
/PfPI3K (cyan)...hVPS34

D2008

D761

c

p110γ ...DHA

Y867

D841

d

```
PfPI3K       GDDLRQDHLVIQIIYVMDNIWKRYGLDLKMTLYRVL
PI3K-C2α     GEDLRQDMLALQMIKIMDKIWLKEGLDLRMVIFKCL
PI3K-C2β     GDDLRQDMLTLQMIRIMSKIWVQEGLDMRMVIFRCF
P110δ        GDDLRQDMLTLQMIQLMDVLWKQEGLDLRMTPYGCL
P110γ        GDDLRQDMLILQILRIMESIWETESLDLCLLPYGCI
Human Vps34  GDDLRQDQLILQIISLMDKLLRKENLDLKLTPYKVL
```

Extended Data Figure 3 | Structural analyses of PfPI3K and human PI3K, with and without artemisinin derived compounds a Snapshots from 20 ns MD simulations and chemical structure of artelinic acid (top right), artemisone (bottom left) and deoxyartemisinin (bottom right) bound to PfPI3K. Snapshot from MD simulation of DHA bound to PfPI3K is shown (at the top left) for reference. Due to the lack of the hydroxyl group in artelinic acid and artemisone, no interactions involving D1889 and Y1915 were observed. Instead, the carboxylic group of artelinic acid forms a salt bridge with R1368 and the sulfone moiety of artemisone interacts with R1368. Deoxyartemisinin lacks hydrogen bond donors, so D1889 is not involved in protein-ligand interaction. Instead, the carbonyl oxygen of deoxyartemisinin forms a hydrogen bond with Y1915. Boxed inset shows the change in the overall shape due to the replacement of the endoperoxide with an ether bridge in deoxyartemisinin (green) and compared to DHA (red), thus leading to loss of the shape complementarity and reorientation in the binding site. As shown in Fig. 1a, c, deoxyartemisinin failed to block PI3P production, in transgenic 3D7 parasites expressing SS-EEA1^{WT}-mCherry (red) or purified PfPI3K. Thus, modelling studies explain a lack of effect of deoxyartemisinin. They would also suggest lack of effect of artelinic acid and artemisone compounds on PfPI3K. It should be noted the target of these compounds is unknown and it remains unclear whether they would share a ring stage target with DHA important for clinical resistance to artemisinins. **b,** Snapshots of MD simulation of DHA–human VPS34 complex. The initial coordinates of DHA–human VPS34 complex was obtained by the overlay of DHA…PfPI3K model (Extended Data Fig. 2b) with the human VPS34 crystal structure (PDB: 3IHY). The obtained complex was then subjected to 20 ns MD simulations. As can be seen, DHA does not interact with D644 and Y670, which correspond to D1889 and Y1915 in PfPI3K, respectively. Instead, the hydroxyl group of DHA now interacts with D761 (left). Snapshot at the right shows an overlay of DHA…human VPS34 complex with the PfPI3K model. In human VPS34 (grey), the loop that contains the D761 residue bends down to interact with DHA. In contrast, D2008 of PfPI3K (cyan), which is in the same position as D761 in human VPS34, has a different orientation which makes it unable to interact with DHA. **c,** Both D1889 and Y1915 in PfPI3K are conserved in human PI3K p110γ. 20 ns MD simulations of the DHA …p110γ complex (PDB code 4ANV), reveal that although the hydrogen bond between DHA and D841 still remains, the interaction between DHA and Y867 is lost. This suggests a weaker binding of DHA to p110γ compared to PfPI3K. The same argument can be made of p110δ based on sequence homology, but no crystal structure of human p110δ is available. **d,** Partial sequence alignment of PfPI3K, with human PI3K-C2α, PI3K-C2β, p110δ, p110γ and VPS34. The key residues of PfPI3K (D1889 and Y1915) that interact with DHA and similar positions in the human kinases are coloured red. While D1889 in PfPI3K is conserved in both PI3K-C2α and PI3K-C2β, Y1915 in PfPI3K is replaced by F1172 and F1117 in PI3K-C2α and PI3K-C2β, respectively. Therefore, the hydrogen bond of the hydroxyl group in Y1915 suggested to be crucial in the computational analysis of PfPI3K cannot be formed with phenylalanine (F). As far as we are aware, no crystal structures of PI3K-C2α and PI3K-C2β have been reported. In p110δ, p110γ and VPS34, a proline N-terminal to Y1915 is conserved. It is likely the reason for the different position and low flexibility of this loop observed in the MD simulations of p110δ, and VPS34.
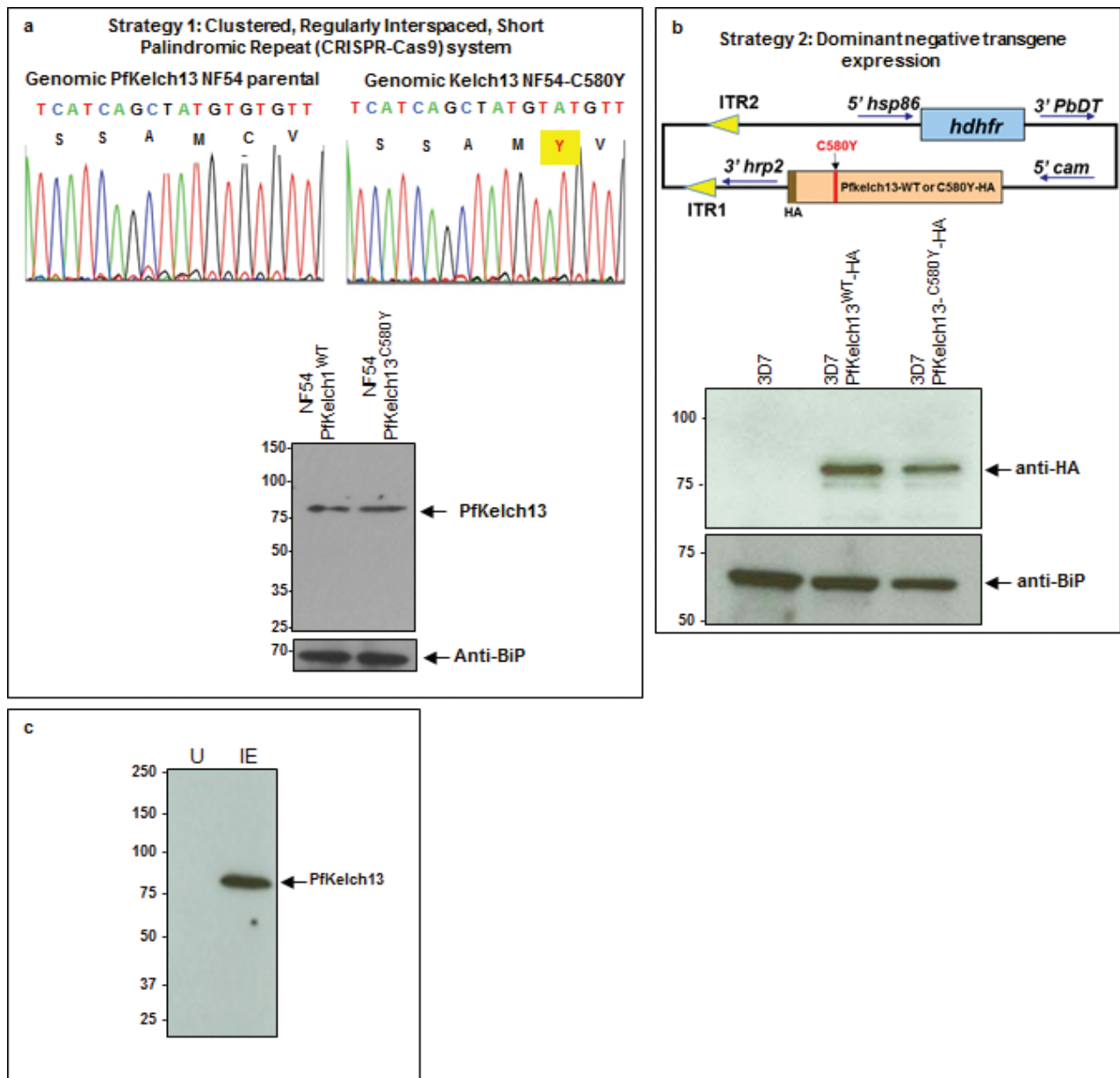
Extended Data Figure 4 | Transgenic *P. falciparum* expressing PfKelch13^C580Y mutation in two different strains of *P. falciparum* using distinct approaches and markers. a, Strategy 1: CRISPR–Cas9 used to introduce a single point mutation in the *P. falciparum* NF54 genome as described elsewhere[20]. Both parent and mutant strains were sequenced to ensure that the only change detected was PfKelch13^C580Y. Western blots using anti-PfKelch13 antibodies confirm that mutation does not change the level of PfKelch13 protein. **b**, Strategy 2 expresses a dominant negative PfKelch13^C580Y and wild-type PfKelch13^WT tagged with HA, and driven by the constitutive *cam* promoter in *P. falciparum* 3D7. Western blots confirm that anti-HA antibodies detected tagged forms of PfKelch13 in transfected but not non-transfected 3D7. Comparable levels of wild-type and mutant transgenes are expressed in resulting two strains of parasites (as shown). In **a** and **b**, BiP, an ER marker serves as a parasite loading control in western blots. Molecular weight standards (in kDa) are as indicated. **c**, Western blot indicate that antibodies to PfKelch13 specifically recognize a ~83 kDa protein in infected (IE) but not uninfected erythrocytes (U). Molecular weight standards (in kDa) are as indicated. In **a–c**, data are representative of three experimental replicates; raw data are in Supplementary Data 2.

a

700 nM DHA (or mock) → DHA wash off → Make smears, count

Invasion

0–3 h Rings — 6–9 h Rings — Culture for 66 h — 72–75 h

$$\text{RSA (Ring- survival assay)} = \frac{\%\ \text{Parasitemia (DHA treatment)}}{\%\ \text{Parasitemia (Mock treatment)}} \times 100$$

b

Parasite strains	PI3P (pmoles/ 6 × 10^7 IE)	RSA	PfKelch13 mutation	PfPI3K polymorphism (I682T, Q431K, Y1330C)
ANL-1 (S)	24.4	0	None	None
ANL-7 (S)*	32.7	1.8	R539T	Q431K
ANL-2 (R)	78.7	17	C580Y	None
ANL-4 (R)	40	6	C580Y	Q431K
ANL-9 (R)	59.5	15	R539T	Q431K
NF54-PfKelch13^WT	22	1	None	None
NF54-PfKelch13^C580Y	59	13.3	C580Y	None
P. falciparum 3D7	19.2	0.5	None	None
3D7-PfKelch13^WT-HA	23	1	None	None
3D7-PfKelch13^C580Y-HA**	49	9	C580Y	None
Pf^VPS34-myc1	54.2	13.1	None	None
Pf^VPS34-myc2	50.2	12.4	None	None
Pf^VPS34-MUTANT-myc	20.0	1.2	None	None
EEA1^R1374A-mCherry (Transfection control)	22.4	1.1	None	None
(3D7) Pf^VPS34-myc1: PfKelch13^WT-HA	42	9.2	None	None
**(3D7) Pf^VPS34-myc1: PfKelch13^C580Y-HA	65	13.3	C580Y	None

* Sanger sequencing revealed that ANL7 was a mixture explaining the presence of the R539T mutation and RSA value of 1.8. ANL7 was therefore considered an anomalous strain not included in subsequent
* *These strains express dominant negative PfKelch13^C580Y as a transgene but also contain endogenous PfKelch13^WT.
For sensitive strains RSA <1. For resistant strains, RSA SE 0.5. For PI3P, SE 0.5

c Vector Schematic for VPS34-myc1, 2 or VPS34-MUTANT

ITR2 — 5' hsp86 — hdhfr or bsd — 3' PbDT
3' hrp2 — ± myc — VPS34 or mutant — 5' cam
ITR1

VPS-myc1 : bsd as selection marker
VPS-myc2/MUTANT : hdhfr as selection marker

d

VPS34-myc | Exp-2 | Merge | + Brightfield

VPS34-myc1
VPS34-myc2

e Vector Schematic for SS-EEA1^R1374A-mCherry

ITR2 — 5' hsp86 — bsd — 3' PbDT
3' hrp2 — ± myc — eea1^R1374A — 5' cam
ITR1

U | IE
50
37 ← SS-EEA1^R1374A-mCherry
25

anti-mCherry

f

Schematic strategy for generating Pf^VPS34-myc1: PfKelch13^WT-HA and Pf^VPS34-myc1: PfKelch13^C580Y-HA parasites

Pf^VPS-myc1 parasites : bsd resistant (generated using strategy described in b)

Transfect with

ITR2 — 5' hsp86 — hdhfr — 3' PbDT
C580Y
3' hrp2 — Pfkelch13-WT or C580Y-HA — 5' cam
ITR1 — HA

Select with WR99210

Pf^VPS34-myc1: PfKelch13WT-HA
and
Pf^VPS34-myc1: PfKelch13C580Y-HA parasites

100
75 ← anti-HA
100
75 ← anti-myc

Pf^VPS-myc1 | Pf^VPS-myc1: PfKelch13^WT-HA | Pf^VPS-myc1: PfKelch13^C580Y-HA

Extended Data Figure 5 | An _in vitro_ ring-stage survival assay (RSA) level is associated with PI3P elevation in the presence and absence of Kelch mutations. **a**, Schematic of the RSA assay[22]. **b**, Summary of RSA, PI3P, PfKelch13 mutation and PfPI3K polymorphisms in clinical and laboratory strains used in this study. Means are shown for three experimental replicates. RSA s.d. <0.5; PI3P s.d. <8 (as shown by error bars in Fig. 3c). **c, d**, Construction of 3D7 transgenic parasites expressing myc- tagged forms of human VPS34, the mammalian orthologue of PfPI3K. The resulting Pf$^{VPS34-myc1/2}$ express the _VPS34_ transgene in ring stage parasites, as detected in indirect immunofluorescence assays (Pf$^{VPS34-MUTANT-myc}$ data not shown in IFA). Exp-2 (red) marks the parasite boundary. Blue, parasite nucleus, P, parasite, E, erythrocyte; scale bar, 5 μm. Fluorescent and merged images are

shown. Whole genome Illumina sequencing indicated VPS34-myc1 was inserted into PF3D7_1363900 and VPS34-myc2 is inserted in PF3D7_0718800. Both PF3D7_1363900 and PF3D7_0718800 encode for unknown function and neither has been reported in GWAS on artemisinin resistance. **e**, Construction of the strain 3D7-SS-EEA1^{R1374A}-mCherry (used as transfection control) and western blot to confirm expression of SS- EEA1^{R1374A}-mCherry (arrow). Molecular weight standards (in kDa) are as indicated. **f**, Construction of Pf$^{VPS34-myc1}$: PfKelch13^{WT}-HA and Pf$^{VPS34-myc1}$: PfKelch13^{C580Y}-HA parasites and western blots confirming expression of the transgenes. Raw data for western blots in **e-f** are shown in Supplementary Data 2. Molecular weight standards (in kDa) are as indicated. In **d–f**, data are representative of three experimental replicates.
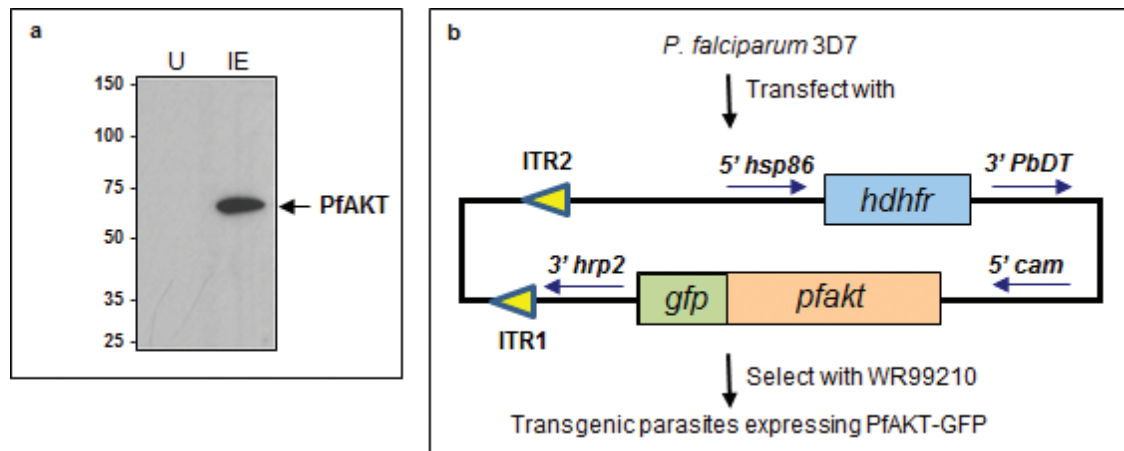
Extended Data Figure 6 | Evidence that the PI3K–AKT pathway is responsive to DHA. **a**, Western blot shows specificity of anti-PfAKT antibody that recognizes a band of expected size in infected (IE) but not uninfected (U) erythrocytes (1×10^7 total cells loaded in each lane). Raw data are in Supplementary Data 2. Data are representative of three experimental replicates. **b**, Strategy and construct for generating transgenic 3D7 parasites expressing PfAKT–GFP.

Extended Data Table 1 | Percentage inhibition of mammalian kinases by 10 μM DHA

Kinase	% Inh	Kinase	% Inh	Kinase	% Inh	Kinase	% Inh
ABL1 T315I	-2	FGFR3 K650E	-2	LCK	1	PRKACA (PKA)	2
ALK	-1	FGFR4	4	MAP3K8 (COT)	3	PRKCA (PKC alpha)	-2
AKT1 (PKB alpha)	3	FRAP1 (mTOR)	3	MAPK1 (ERK2)	2	PRKCQ (PKC theta)	-3
AURKA (Aurora A)	3	GSK3B (GSK3 beta)	0	MAPK14 (p38 alpha)	1	RET	3
AXL	-1	IGF1R	2	MAPKAPK2	5	ROCK2	1
BTK	-1	INSR	4	MET (cMet)	2	SYK	-2
CDK2/cyclin A	1	JAK1	0	PDGFRA V561D	-1	TYK2	5
EGFR (ErbB1)	-1	JAK2	1	PDK1 Direct	-1	ZAP70	1
EPHA4	2	JAK3	-4	PI4KB (PI4K beta)	-13		
EPHB4	1	KDR (VEGFR2)	0	PKN1 (PRK1)	0		
ERBB2 (HER2)	-4	KIT	-1	PKN2 (PRK2)	-3		

s.d. <0.5.

Extended Data Table 2 | Primers used for cloning

Primer	Sequence (5'-3')
PfKelch13AvrIIF	AGATCTCCTAGGATGGAAGGAGAAAAAGTAAAAACAAAAGCAAATAG
PfKelch13HAXhoIR	GGATCCCTCGAGTTAAGCATAATCTGGAACATCATATGGATAAGCAGCTGCTATATTTGCTATTAAA ACGGAGTGACCAAATCTGG
C580Yreverse	TTATTATCAAAAGCAACATACATAGCTGATGATCTAGG
C580Yforward	CCTAGATCATCAGCTATGTATGTTGCTTTTGATAATAA
HsVps34AvrIIF	GGATCCCCTAGGATGGGGGAAGCAGAGAAGTTTCACTACATC
HsVps34mycXhoIR	GGATCCCTCGAGTTATAAATCTTCTTCAGATATTAATTTTTGTTCAGCTGCTGCTTTTCTCCAGTACT GGGCAAACTTGTGAATC
Vps34-742AAA745F	GGAGTTGGAGCTGCAGCACTGGATAACCTTTTGCTAACAAAAACAGGCAAACTC
Vps34-742AAA745R	GTTATCCAGTGCTGCAGCTCCAACTCCAAGTATATAGGTGATCACGCAATATCC
AKTAvrII_F	GTCCTAGGATGAATCATATAAACTTTTTCAAATG
AKTBglII_R	GTAAGATCTTTTTTGTTGACCTGAAAATTCATAATAG
Kelch13_F	ATGGAAGGAGAAAAAGTAAAAACAAAAG
Kelch13_R	TATATTTGCTATTAAAACGGAGTGACC
Kelch13-1	ATGGAAGGAGAAAAAGTAAAAAC
Kelch13-2	GATATGAGTGTATTAGATTCGAAC
Kelch13-3	GATACTTATGAAAAGAAAATTATTG
Kelch13-4	GAAAAAGAAAAATATTATCAAGAA
Kelch13-5	CAACAATGCTGGCGTATGTGTACACC
Kelch13-6	GATAATAAAATTTATGTCATTGGTGG
PfPI3K-I682T-F	CCTCATAGTACGAATAATACAAATCTCCAATCC
PfPI3K-I682T-R	ATATATTAATTCTTTATTAAAATTAAGGAAATTCAAAATCATGTATCC
PfPI3K-Q431K-F	AACCTTTTCATATGATAACAATCCAAACAATATC
PfPI3K-Q431K-R	TTTTCTTTTTATATGTCTTATAATTGTAACGACTGG
PfPI3K-Y1330C-F	CCAAGAATAAAACAGACAGAATACAAAATAAAACC
PfPI3K-Y1330C-F	CTCATAGGAGTATCCAGCGAATTTTCC

Single–dose attenuated Vesiculovax vaccines protect primates against Ebola Makona virus

Chad E. Mire[1,2]*, Demetrius Matassov[3]*, Joan B. Geisbert[1,2], Theresa E. Latham[3], Krystle N. Agans[1,2], Rong Xu[4], Ayuko Ota–Setlik[4], Michael A. Egan[4], Karla A. Fenton[1,2], David K. Clarke[3], John H. Eldridge[3,4] & Thomas W. Geisbert[1,2]

The family *Filoviridae* contains three genera, Ebolavirus, Marburgvirus, and Cuevavirus[1]. Some members of the genus, including Zaire ebolavirus (ZEBOV), can cause lethal haemorrhagic fever in humans. During 2014 an unprecedented ZEBOV outbreak occurred in West Africa and is still ongoing, resulting in over 10,000 deaths, and causing global concern of uncontrolled disease. To meet this challenge a rapid-acting vaccine is needed. Many vaccine approaches have shown promise in being able to protect nonhuman primates against ZEBOV[2]. In response to the current ZEBOV outbreak several of these vaccines have been fast tracked for human use. However, it is not known whether any of these vaccines can provide protection against the new outbreak Makona strain of ZEBOV. One of these approaches is a first-generation recombinant vesicular stomatitis virus (rVSV)-based vaccine expressing the ZEBOV glycoprotein (GP) (rVSV/ZEBOV). To address safety concerns associated with this vector, we developed two candidate, further-attenuated rVSV/ZEBOV vaccines. Both attenuated vaccines produced an approximately tenfold lower vaccine-associated viraemia compared to the first-generation vaccine and both provided complete, single-dose protection of macaques from lethal challenge with the Makona outbreak strain of ZEBOV.

Since discovery of the virus in 1976, outbreaks of ZEBOV have been detected sporadically in Africa. With increasing population growth the frequency of human contact with natural virus reservoirs[3] is likely to rise, potentially leading to more catastrophic outbreaks such as the current epidemic in West Africa, thus increasing the need for effective antiviral strategies. A highly effective countermeasure would be a preventive vaccine that can be simply and widely administered to people in regions of virus zoonosis and provide a 'blanket immunity' curtailing any future outbreaks. Also important will be the ability to rapidly combat deliberate misuse of these deadly viruses. Therefore, a preventive vaccine should ideally confer rapid, single-dose protection.

There are currently no licensed filovirus vaccines or post-exposure treatments available for human use. However, there are at least ten different vaccine approaches that have shown the potential to protect nonhuman primates (NHPs) from lethal ZEBOV infection, including platforms based on recombinant adenovirus serotype 5 (rAd5) vectors, combined DNA/rAd5 vectors, combined rAd serotype 26 and 35 vectors, recombinant chimpanzee adenovirus serotype 3 (rChAd3) vectors, combined rChAd3 and modified vaccinia Ankara (MVA) vectors, virus-like particles (VLPs), alphavirus replicons, recombinant human parainfluenza virus 3 (rHPIV3), rabies virus, and recombinant vesicular stomatitis virus (rVSV)[2]. Of the vaccines advancing to phase I trials, the rChAd3 and rVSV vectored vaccines have shown success in single-dose protection of NHPs against ZEBOV challenge; with the caveat that the rChAd3/ZEBOV vaccine requires a boost with an MVA/ZEBOV vector for protection past 6 months[4]. Also, NHPs inoculated with the rChAd3/ZEBOV vaccine were challenged with a ZEBOV seed stock containing a large virus population encoding 8 uridines (U) at a critical transcription editing site in the GP gene[4]. This specific genetic feature typically arises following prolonged passage of ZEBOV in Vero E6 cells and results in higher levels of expression of full-length GP. In contrast, low-passage ZEBOV isolates retain 7U at the GP editing site, resulting in higher levels of secreted GP (sGP) expression, which is associated with greater viral virulence[5–7]. Importantly, studies have shown that rAd-based ZEBOV vaccines that completely protect NHPs against ZEBOV stocks containing high populations of 8U virus are not able to completely protect vaccinated macaques challenged with ZEBOV stocks containing high populations of 7U virus[8].

The first generation rVSV/ZEBOV vaccine that replaces the VSV glycoprotein G with the ZEBOV GP (rVSV/ZEBOVΔG), originally developed by Drs Feldmann and Geisbert[9] and currently licensed by Merck, has demonstrated solid single-dose NHP protection against a low-passage 7U ZEBOV stock[8]. The rVSV/ZEBOVΔG vector has also protected 50% of NHPs when administered shortly after ZEBOV challenge[10], and has demonstrated safety in a NHP neurovirulence model[11]. However, there is a robust post-vaccination viraemia in macaques and a recent phase I trial of the rVSV/ZEBOVΔG vaccine in Geneva was halted due to temporary joint pain in some patients. The level of vaccine-associated viraemia and frequency of adverse events will be more fully documented as data from ongoing phase 3 trials become available for this vector; but the early observation suggest that a further-attenuated rVSV vector may be more desirable for widespread administration in endemic regions of Africa.

To address this possible safety concern we have developed and tested two further-attenuated rVSV/ZEBOV vaccine candidates for efficacy. One of these vaccines is based on an rVSV vector that has advanced through clinical evaluation. It was attenuated by translocating the VSV nucleoprotein (N) gene from position 1 to position 4 in the genome (N4) and truncating the cytoplasmic tail (CT) of the VSV G protein from 29 to 1 amino acids (CT1)[12]. This rVSVN4CT1 vector was modified to maximally express HIV-1 gag from position 1 in the genome (rVSVN4CT1gag1) by positioning the *gag* gene immediately adjacent to the single strong 3′ VSV transcription promoter. The rVSVN4CT1gag1 vector has demonstrated safety in mouse and NHP neurovirulence studies[12,13], and replication is restricted to the IM inoculation site and draining lymph node following vaccination of mice[14]. The rVSVN4CT1gag1 vector has demonstrated safety and immunogenicity in two phase I clinical trials (HVTN 090 and HVTN 087: http://clinicaltrials.gov/) and no post-vaccination viraemia was detected in urine, saliva, and blood of vaccine recipients. The rVSVN4CT1GP1 vector described here (Fig. 1a, N4) is analogous in design to that of the rVSVN4CT1gag1 vaccine and expresses ZEBOV GP from genome position 1. The other attenuated rVSV/ZEBOV
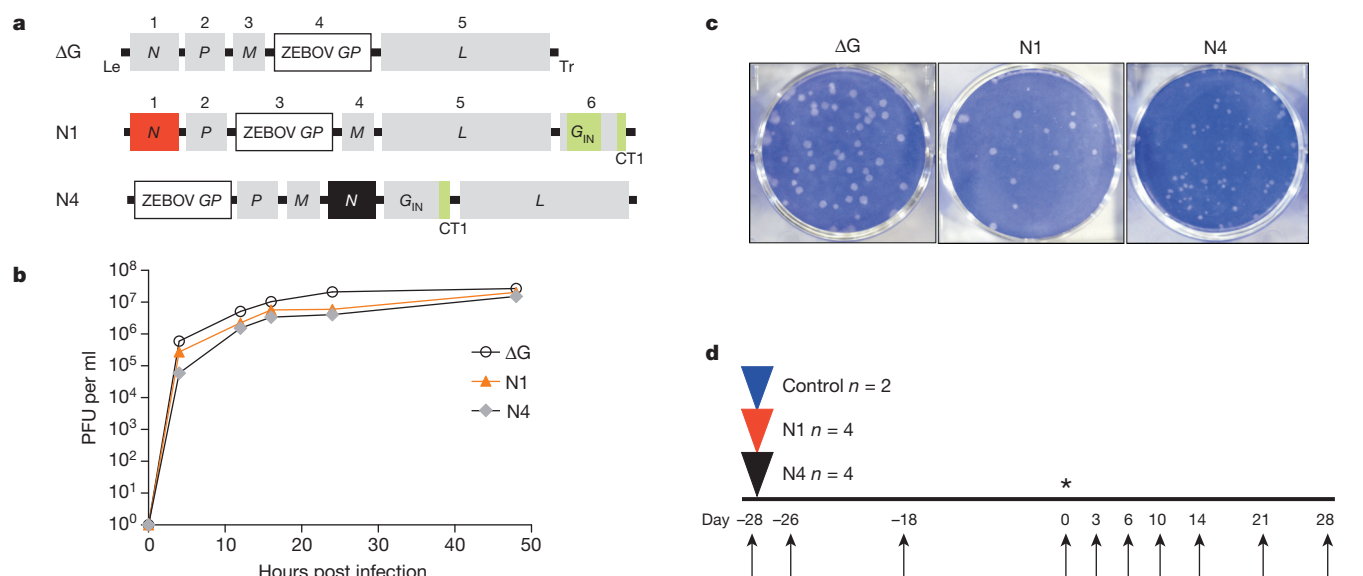
Figure 1 | rVSV/ZEBOV vector design, growth kinetics and vaccine study strategy. a, Genome organization comparing ZEBOV GP (Mayinga strain)-expressing rVSV vectors as described in methods. The rVSV/ZEBOVΔG (ΔG) vector had the natural VSV *G* gene replaced with the ZEBOV *GP* at position 4 within the genome. rVSVN1CT1GP3 (N1) vector retained the position of VSV *N* in position 1 (orange box), insertion of ZEBOV *GP* at position 3 and a truncated form of VSV *G* containing the CT1 truncation was inserted at position 6. The rVSVN4CT1GP1 (N4) vector had the insertion of ZEBOV *GP* in position 1, attenuating *N* gene translocation (N4) (black box) and truncated G protein cytoplasmic tail (CT1). Numbers above vector constructs designate genome positions. Virus leader (Le), trailer (Tr), and intergenic regions are shown in black. Shaded regions represent deleted amino acid regions. **b**, Single-cycle growth kinetics comparing the ΔG, N1, and N4 vectors depicted in **a**. Data shown are mean ± s.d. from two biological replicates titrated by plaque assay in triplicate. Titre differences between ΔG and N1 vectors were statistically significant at 4 ($P = 0.0001$), 12 ($P = 0.0055$), and 24 h post infection ($P = 0.0001$). Likewise, ΔG and N4 vector titres were significantly different at 4 ($P = 0.0001$), 12 ($P = 0.0005$), 24 ($P = 0.0001$), and 48 h post infection ($P = 0.0068$). Unpaired *t*-test, $P = 0.05$. **c**, Crystal violet-stained Vero cell monolayers showing plaques generated by the ΔG, N1, and N4 vectors at 48 h post infection. **d**, Flow chart showing the day of vaccination (triangles), days of sampling (arrows), day of challenge (*). Blue triangle, unvaccinated cohort; orange triangle, N1-vaccinated cohort; black triangle, N4-vaccinated cohort.

vaccine described here (rVSVN1CT1GP3), expressing a truncated form of VSV G, was designed to be of intermediate attenuation between rVSVN4CT1GP1 and the first generation rVSV/ZEBOVΔG vaccine (Fig. 1a, N1). Both attenuated rVSV/ZEBOV vectors express GP from the ZEBOV Mayinga strain, as do most other candidate ZEBOV vaccines currently under evaluation. Sequence homology between GPs from the new West African Makona strains analysed to date and the 1976 Mayinga strain is approximately 97%. Although this difference is not likely to affect the protective efficacy of the current ZEBOV vaccines against the heterologous West African strains, it is possible that small changes in sequence could lead to reduced efficacy of a vaccine[15]. It is well established that small variations in sequence and even single amino acid changes in sequence for other viruses including influenza, respiratory syncytial virus, polio, equine infectious anaemia virus, and SIV can reduce vaccine efficacy. Here we assessed the ability of our newly developed next-generation rVSV-based vaccines expressing ZEBOV Mayinga GP to protect against heterologous challenge with the new outbreak Makona strain of ZEBOV in cynomolgus monkeys.

Table 1 | Clinical findings for NHPs challenged with ZEBOV-Guinea

Animal	Vaccine	Day −26*	PRNT50†	Clinical signs observed‡	Final outcome
129	N/A	−	0/0	Fever (6), anorexia (5–8), depression (6–8), mild rash (6–8), lymphopenia (3, 6), thrombocytopenia (6), ALT→→(6), ALP→→ → (6), AST→→→ (6), GGT→→→ (6), CRP increase (6)	Expired day 8
276	N/A	−	0/0	Fever (6), anorexia (6–7), depression (6–7), mild rash (6–7), thrombocytopenia (6, 10), ALT→(6), ALP→→→ (6), AST→→→ (6), GGT→ (6), CRP increase (6)	Expired day 7
0910078	N1	+	0/40	Ø§	Survived
1001100	N1	−	0/160	CRP increase (6)	Survived
117	N1	−	0/80	Lymphopenia (6), CRP increase (6, 10)	Survived
0907095	N1	−	0/160	Lymphopenia (6, 10), CRP increase (6, 10), ALT→→(6), ALP→(6), AST→→(6)	Survived
0807174	N4	+	0/160	Lymphopenia (6), CRP increase (6, 10)	Survived
0901014	N4	−	0/80	Ø	Survived
119	N4	+	0/80	Ø	Survived
0811013	N4	+	0/20	Ø	Survived

* rVSV viraemia 2 days post vaccination. −, below limit of detection (25 PFU per ml); +, up to 3×10^2 PFU per ml.
† 50% plaque reduction neutralization titre at day of challenge and terminal day presented as day of challenge/terminal day.
‡ Days after ZEBOV challenge are in parentheses. Fever is defined as a temperature more than 1.4 °C above baseline or at least 0.9 °C above baseline and ≥ 39.7 °C. Lymphopenia and thrombocytopenia are defined by a ≥ 35% drop in numbers of lymphocytes and platelets, respectively. ALT, alanine aminotransferase; ALP, alkaline phosphatase; AST, aspartate aminotransferase; GGT, gamma glutamyltransferase. CRP, C-reactive protein: two- to threefold increase,→; 4- to fivefold increase, →→; > 5 fold increase, →→→.
§ No symptoms observed.

Results from an *in vitro* growth kinetics study (Fig. 1b) indicate an approximate tenfold reduction in growth rate early in infection for rVSVN4CT1GP1 relative to rVSV/ZEBOVΔG. Also noted during virus plaque assay were the larger more rapidly forming plaques generated by rVSV/ZEBOVΔG compared to rVSVN4CT1GP1, with rVSVN1CT1GP3 showing intermediate growth and plaque size (Fig. 1c).

We next tested if the further-attenuated rVSV/ZEBOV vaccines could provide NHPs with single-dose protection against challenge with ZEBOV isolated from the current outbreak in Guinea[16]. Groups of four cynomolgus macaques were inoculated intramuscularly with 2×10^7 plaque-forming units (PFU) of either rVSVN4CT1GP1 or rVSVN1CT1GP3; a group of two control macaques were unvaccinated (Fig. 1d, arrow heads). None of the macaques showed any sign of illness or distress following vaccine administration. Consistent with the statistically significant growth differences between rVSV/ZEBOVΔG and the more attenuated vectors seen during *in vitro* growth kinetics studies, levels of both attenuated vaccine viruses detected in the blood of vaccinated macaques (500 PFU per ml) were 10- to 50-fold lower than those detected for the more replication competent rVSV/ZEBOVΔG[9] (Table 1, day −26). The ZEBOV GP-specific humoral immune response was assessed for all animals before vaccination (Fig. 2a, −28) and after vaccination (Fig. 2a, −18 and 0) by IgG capture ELISA and neutralizing antibody titres (Table 1, plaque 50% reduction neutralization test (PRNT)$_{50}$). Results showed neutralizing titres at terminal days for vaccinated cohorts and detectable circulating levels of anti-ZEBOV GP IgG for both vaccine cohorts after vaccination and before challenge with no detectable levels for the unvaccinated control animals (Fig. 2a). A cell-mediated immune response was also detected in all vaccinated animals by ZEBOV GP-specific interferon gamma (IFN-γ) ELISpot assay 10 days after vaccination (Extended Data Fig. 1a and b).

The eight vaccinated and two unvaccinated control macaques were challenged by intramuscular injection with 1,000 PFU of a low passage 100% 7U Makona strain stock of ZEBOV[16] 28 days after the single injection vaccination (Fig. 1d, asterisk). None of the animals

vaccinated with either of the two further-attenuated rVSV/ZEBOV vectors showed any severe signs of illness following challenge with ZEBOV (Table 1), whereas the two unvaccinated control macaques succumbed to disease on days 7 and 8 (Fig. 2b). Circulating infectious ZEBOV was isolated from both of the unvaccinated control macaques on days 3 and 6 post challenge (Fig. 2c, blue) but no circulating infectious ZEBOV could be detected in any of the vaccinated animals. Examination of tissues by immunohistochemistry showed abundant ZEBOV antigen in tissues of the unvaccinated control animals (129 and 276) (Fig. 3a–d) whereas ZEBOV antigen was not detected in tissues of the macaques vaccinated with rVSVN1CT1GP3 (1001100) or rVSVN4CT1GP1 (0807174) (Fig. 3e–h).

Here we show protection against a new West African Makona strain of ZEBOV using a novel filovirus vaccine platform. The large reduction in vaccine-associated viraemia indicates a significant increase of *in vivo* attenuation for these next-generation rVSV/ZEBOV vaccine vectors, which should translate into greater safety and reduced adverse

Figure 2 | N1 and N4 vaccination results in circulating anti-ZEBOV GP IgG and protection in cynomolgus macaques. a, Reciprocal endpoint dilution titres for circulating IgG against ZEBOV GP for control (blue), N1 (orange), and N4 cohorts (black-grey) on day of vaccination (−28), 10 days post vaccination (−18), and on day of challenge (0). Red dashed line depicts limit of detection for ELISA assay. Error bars represent s.e.m. **b**, Kaplan–Meier survival curve for each cohort post ZEBOV challenge. **c**, Circulating infectious virus load displayed as plaque forming units per ml. Data shown are from individual animals. Lower limit of detection is 25 PFU per ml.

Figure 3 | Comparison of ZEBOV antigen in tissues of cynomolgus macaques either vaccinated or unvaccinated. a, c, Liver, diffuse cytoplasmic immunolabelling (brown) of sinusoidal lining cells in both ZEBOV-infected control animals. **b, d**, Spleen, diffuse cytoplasmic immunolabelling of dendriform mononuclear cells in the red and white pulp of ZEBOV-infected control animals. **e, f**, Liver and spleen, respectively, with a lack of immunolabelling from N1 cohort animal 0910078. **g, h**, Liver and spleen, respectively, with a lack of immunolabelling from N4 cohort animal 0807174.

events in humans. Importantly, single-dose vaccination of NHPs with highly attenuated forms of rVSV expressing ZEBOV Mayinga GP provides complete protection from heterologous challenge with a highly virulent 7U ZEBOV isolated early during the current West African outbreak[16]. ZEBOV genome sequencing from cases later during the West Africa outbreak has revealed little drift in the GP gene[17,18], suggesting that this vaccine platform could also be efficacious against currently circulating ZEBOV. These findings pave the way for the identification and manufacture of safer, single-dose, high efficacy vaccine(s) to combat current and future filovirus outbreaks in Africa and their potential use as biological weapons.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Feldman, H., Sanchez, A. & Geisbert, T. W. in *Fields Virology* Ch. 32 (Lippincott Williams and Wilkins, 2013).
2. Mire, C. E. & Geisbert, T. W. in *Biology and Pathology of Rhabdo- and Filoviruses* Ch. 22 (World Scientific Publishing, 2014).
3. Leroy, E. M. *et al.* Fruit bats as reservoirs of Ebola virus. *Nature* **438**, 575–576 (2005).
4. Stanley, D. A. *et al.* Chimpanzee adenovirus vaccine generates acute and durable protective immunity against ebolavirus challenge. *Nature Med.* **20**, 1126–1129 (2014).
5. Volchkova, V. A., Dolnik, O., Martinez, M. J., Reynard, O. & Volchkov, V. E. Genomic RNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. *J. Infect. Dis.* **204** (Suppl 3), S941–S946 (2011).
6. Kugelman, J. R. *et al.* Ebola virus genome plasticity as a marker of its passaging history: a comparison of *in vitro* passaging to non-human primate infection. *PLoS ONE* **7**, e50316 (2012).
7. Mohan, G. S., Li, W., Ye, L., Compans, R. W. & Yang, C. Antigenic subversion: a novel mechanism of host immune evasion by Ebola virus. *PLoS Pathog.* **8**, e1003065 (2012).
8. Hirschberg, R. *et al.* Challenges, progress, and opportunities: proceedings of the filovirus medical countermeasures workshop. *Viruses* **6**, 2673–2697 (2014).
9. Jones, S. M. *et al.* Live attenuated recombinant vaccine protects nonhuman primates against Ebola and Marburg viruses. *Nature Med.* **11**, 786–790 (2005).
10. Feldmann, H. *et al.* Effective post-exposure treatment of Ebola infection. *PLoS Pathog.* **3**, e2 (2007).
11. Mire, C. E. *et al.* Recombinant vesicular stomatitis virus vaccine vectors expressing filovirus glycoproteins lack neurovirulence in nonhuman primates. *PLoS Negl. Trop. Dis.* **6**, e1567 (2012).
12. Clarke, D. K. *et al.* Synergistic attenuation of vesicular stomatitis virus by combination of specific G gene truncations and N gene translocations. *J. Virol.* **81**, 2056–2064 (2007).
13. Clarke, D. K. *et al.* Neurovirulence and immunogenicity of attenuated recombinant vesicular stomatitis viruses in nonhuman primates. *J. Virol.* **88**, 6690–6701 (2014).
14. Johnson, J. E. *et al. In vivo* biodistribution of a highly attenuated recombinant vesicular stomatitis virus expressing HIV-1 Gag following intramuscular, intranasal, or intravenous inoculation. *Vaccine* **27**, 2930–2939 (2009).
15. Kugelman, J. R. *et al.* Evaluation of the potential impact of Ebola virus genomic drift on the efficacy of sequence-based candidate therapeutics. *MBio* **6**, e02227–14 (2015).
16. Baize, S. *et al.* Emergence of Zaire Ebola virus disease in guinea. *N. Engl. J. Med.* **371**, 1418–1425 (2014).
17. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
18. Hoenen, T. *et al.* Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science* http://dx.doi.org/10.1126/science.aaa5646 (2015).

Author Contributions D.K.C., D.M. and T.E.L. designed the vaccine vectors and did preparative work. J.H.E. M.A.E., A.O.-S. and R.X. designed, conducted, and analysed the *in vitro* vaccine characterization studies. C.E.M., J.H.E., and T.W.G. conceived and designed the NHP study. C.E.M., J.B.G. and T.W.G. performed the NHP vaccination and challenge experiments, and conducted clinical observations of the animals. J.B.G. and K.N.A. performed the clinical pathology assays. J.B.G. performed the ZEBOV infectivity assays. C.E.M., D.M., J.B.G., K.N.A., M.A.E., K.A.F., D.K.C., J.H.E. and T.W.G. analysed the data. K.A.F. performed histologic and immunohistochemical analysis of the data. C.E.M., D.M., D.K.C. and T.W.G. wrote the paper. All authors had access to all of the data and approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to T.W.G. (twgeisbe@utmb.edu)

METHODS

No statistical methods were used to predetermine sample size.

Generation of N4 and N1 ZEBOV vectors. As described previously[12,19] an rVSV$_{IN}$N4CT1gag1 vector (Indian serotype) expressing HIV-1 gag was used as the backbone for generating the attenuated rVSVN4CT1 vector expressing the Zaire ebolavirus (ZEBOV) glycoprotein (GP). The corresponding rVSV$_{IN}$N4CT1gag1 genomic cDNA was modified by exchanging the gag gene expression cassette via XhoI/NotI restriction sites with an expression cassette encoding a full length ZEBOV GP [1976, Mayinga strain], generating the rVSV$_{IN}$N4CT1-ZEBOVGP1 cDNA (Fig. 1a, N4). The N1 vector was generated by first inserting ZEBOV GP into a VSV-N1ΔG backbone via XhoI/NotI restriction sites within a transcriptional cassette located at position 3 in the genome; followed by the insertion of a PCR fragment containing a portion of VSV L, a modified VSV G CT1 gene and trailer into the N1 genome at position 6 via the HindIII/RsrII sites, generating the rVSV$_{IN}$N1(G CT1)6-ZEBOVGP3 cDNA (Fig. 1a, N1).

The rVSV-ZEBOV vectors were rescued from genomic cDNA as previously described[20]. Rescued virus was plaque purified and amplified on Vero cell monolayers (ATCC, CCL-81). For animal studies, virus vectors were purified from infected BHK-21(ATCC CCL-10) cell supernatants by centrifugation through a 10% sucrose cushion. Purified virus was resuspended in PBS, pH 7.0, mixed with a sucrose phosphate (SP) stabilizer (7 mM K_2HPO_4, 4 mM KH_2PO_4, 218 mM sucrose), snap frozen in ethanol/dry ice and stored at $-80\,°C$ until ready for use.

Growth kinetics study of ΔG, N4 and N1 ZEBOV vectors. Single-step growth curves were performed by adsorbing the N4, N1 and a ΔG control virus to duplicate monolayers of Vero cells (ATCC, CCL-81) in six-well plates at a multiplicity of infection (MOI) of 10 for 15 min at room temperature with continued rocking followed by incubation at $37\,°C$ with 5% CO_2 for 30 min without agitation. The inoculum was aspirated, the cells washed $3\times$ with serum-free Dulbecco's minimal Eagle's medium (DMEM) and then DMEM containing 5% fetal bovine serum (FBS) was added to the plates, which were placed at $32\,°C$ with 5% CO_2. Samples for titration were taken at 4, 8, 12, 16, 24 and 48 h post infection and replaced with the same volume of fresh media. Virus titres were determined in duplicate by plaque assay on Vero cells. Growth curves were performed in triplicate for each virus. Plaque images for each vector were taken at 48 h post infection, after staining with a 1% crystal violet solution. Statistical analysis of rVSV titres were performed using unpaired t-test with a 95% confidence level ($P < 0.05$) with the GraphPad Prism program.

Challenge virus. The ZEBOV Makona strain seed stock originated from serum from a fatal case early during the 2014 outbreak in Guékédou, Guinea[16] (NCBI accession number KJ660347) and was passaged twice in Vero E6 cells (ATCC, CRL-1586). The virus stock was deep sequenced as 100% 7U at the GP editing site in the viral genome (see below).

Deep sequencing. Approximately 1 ml of the ZEBOV Makona strain seed stock was removed from the seed stock vial and placed in 5 ml of TRIzol LS and vortexed three times and allowed to sit for 10 min. The 6 ml were then placed into two separate 3 ml Nunc cryo-vials for removal into the BSL-4. RNA was isolated from the TRIzol LS/sample mixture using Zymo Research Direct-zol RNA mini-prep per manufacturer's instructions. Approximately 150 ng of purified RNA were used to make cDNA using the NuGen Ovation RNA-seq 2.0 kit ultimately for the preparation of the double-stranded DNA library using Encore Ion Torrent library prep kit. Sequencing was performed by the UTMB Molecular Core on the Ion Torrent using 318-v2 deep sequencing chips. Sequence analysis was performed using DNA Star Seqman NGen software based on paired-end analysis of 100 base pairs overlaps.

Vaccination and animal challenge. Ten, healthy, filovirus-naive, adult (~3 to 9.5 kg, 7 female and 3 male), Chinese origin cynomolgus macaques (*Macaca fascicularis*) were randomized with Microsoft Excel into two experiment groups of four animals each and a control group of two animals. Animals in one experimental group were vaccinated by intramuscular injection of approximately 2×10^7 PFU of the rVSVN4CT1GP1vaccine while animals in the other experimental group were vaccinated with approximately 2×10^7 PFU of the VSV-N1CT1 ZEBOVGP vaccine. The two control animals were not vaccinated. Four weeks after the single injection vaccination all ten animals were challenged by intramuscular injection with 1,000 PFU of the ZEBOV Makona strain virus. All animals were given physical exams and blood was collected before vaccination, at day 10 after vaccination, at the time of ZEBOV challenge and on days 3, 6, 10, 14, 21 and 28 after ZEBOV challenge (Fig. 1d, arrows). Animals were monitored daily and scored for disease progression with an internal filovirus scoring protocol approved by the UTMB Institutional Animal Care and Use Committee. The scoring changes measured from baseline included posture/activity level, attitude/behaviour, food and water intake, weight, respiration, and disease manifesta-

tions such as visible rash, haemorrhage, ecchymosis, or flushed skin. A score of ≥ 9 indicated that an animal met criteria for euthanasia. This study was not blinded.

Anti-ZEBOV GP IgG ELISA. Serum collected at indicated time points was tested for immunoglobulin G (IgG) antibodies against ZEBOV. Enzyme-linked immunosorbent assay (ELISA) using recombinant ZEBOV GPdTM purified protein (Integrated BioTherapeutics, Inc.) was used to detect cross-reactive IgG. ZEBOV GPdTM was diluted to an optimal working concentration of 100 ng per well in 0.1 ml carbonate/bicarbonate buffer (carbonate/bicarbonate buffer with azide tablets from Sigma catalogue number 08058-50TAB-F) and used to coat Immulon 2HB flat bottom ELISA plates (Thermo Labsystem catalogue number 3455) for 18 h at 4 °C. Coated plates were blocked (10% FBS + $1\times$ PBS) for at least 2 h. The serum samples were assayed at twofold dilutions starting at a 1:100 dilution in ELISA diluent (1% heat inactivated fetal bovine serum (HI-FBS), $1\times$ PBS, and 0.2% Tween-20). Samples were incubated for 1 h at room temperature, removed, and plates were washed. Wells were then incubated at room temperature for 1 h with anti-monkey IgG conjugated to horseradish peroxidase (Fitzgerald Industries International) at a 1:2,500 dilution. These wells were washed and then incubated with 2,2'-azine-di(3ethylbenzthiazoline-6-sulfonate) peroxidase substrate system (KPL) at room temperature for approximately 10 min. Reaction was stopped with 1% SDS and read for dilution endpoints at 405 nm on a microplate reader (Molecular Devices Emax system). Absorbance values were normalized by subtraction of background $A_{405\,nm}$ from uncoated wells for each serum dilution. Antigen-specific serum IgG end-point titres were defined as the reciprocal of the last normalized serum dilution giving an $A_{405\,nm}$ greater than 0.1.

ZEBOV neutralization assay. Neutralization assays were performed by measuring plaque reduction in a constant virus:serum dilution format as previously described[9]. Briefly, a standard amount of ZEBOV (~100 PFU) was incubated with serial twofold dilutions of the serum sample for 60 min. The mixture was used to inoculate Vero E6 cells (ATCC, CRL-1586) for 60 min. Cells were overlaid with an agar medium, incubated for 7 days, and plaques were counted 48 h after neutral red staining. Endpoint titres were determined by the dilution of serum, which neutralized 50% of the plaques ($PRNT_{50}$).

IFN-γ ELISpot assay. Ninety-six-well flat-bottomed ELISpot plates (Millipore) were coated overnight with a mouse anti-human IFN-γ monoclonal antibody (clone 27; BD-Pharmingen) at a concentration of $10\,\mu g\,ml^{-1}$, after which the plates were washed three times with $1\times$ PBS and then blocked for 2 h with PBS containing 5% heat-inactivated FBS. Heparinized whole blood was collected 10 days after immunization of macaques, and peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by Ficoll-Hypaque density gradient centrifugation, and resuspended in complete R05 culture medium. The isolated macaque PBMCs were washed once with complete R05 culture medium and resuspended in complete R05 culture medium containing either $5\,\mu g\,ml^{-1}$ phytohaemagglutinin mucoprotein (Sigma), peptide pools (15-mers overlapping by 11 amino acids; final peptide concentration, $1\,\mu M$ each) spanning the ZEBOV Mayinga strain GP, or medium alone. The input cell number was 2×10^5 PBMCs per well (2×10^6 PBMCs per ml), and cells were assayed in duplicate wells. Cells were incubated for 18 to 24 h at 37 °C and then removed from the ELISpot plate by first being washed with deionized water and then being washed six times with $1\times$ PBS containing 0.25% Tween 20. Thereafter, plates were treated with a rabbit polyclonal anti-human IFN-γ biotinylated detection antibody (0.65 μg per well; Life Technologies) diluted with $1\times$ PBS containing 1% bovine serum albumin (BSA) and were incubated at 37 °C for 2 h. ELISpot plates were then washed 6 times with $1\times$ PBS containing 0.25% Tween 20, treated with 100 μl per well of streptavidin–horseradish peroxidase conjugate (BD Biosciences) diluted 1:250 with $1\times$ PBS containing 10% FBS and 0.005% Tween 20, and incubated for an additional 1 h at room temperature. Unbound conjugate was removed by rinsing the plate six times with $1\times$ PBS containing 0.25% Tween 20 and three times with $1\times$ PBS. A chromogenic substrate (100 μl per well) (one-step nitroblue tetrazolium/5-bromo-4-chloro-3-indolylphosphate (NBT/BCIP); Pierce) was then added for 3 to 5 min before being rinsed away with water, after which the plates were air dried and the resulting spots counted using an ImmunoSpot reader (CTL Inc.). Peptide-specific IFN-γ ELISpot responses were considered positive if the responses (minus the medium background) were threefold above the medium response and 50 spot-forming cells (SFC) per 10^6 PBMCs. Unpaired t-test analysis of IFN-γ ELISpot data was performed on GraphPad Prism version 5.02 software. Two-tailed P values less than 0.05 indicated that the tests were statistically significant.

Detection of viraemia. Virus titration of the rVSV vaccine vectors and ZEBOV was performed by plaque assay with Vero E6 cells (ATCC, CRL-1586) from cell culture or serum samples as previously described[9]. Briefly, increasing tenfold dilutions of the samples were adsorbed to Vero E6 monolayers in duplicate wells (200 μl); the limit of detection was 25 PFU per ml.

Haematology and serum biochemistry. Total white blood cell counts, white blood cell differentials, red blood cell counts, platelet counts, haematocrit values, total haemoglobin concentrations, mean cell volumes, mean corpuscular volumes, and mean corpuscular haemoglobin concentrations were analysed from blood collected in tubes containing EDTA using a laser based haematologic analyser (Beckman Coulter). Serum samples were tested for concentrations of albumin, amylase, alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, gamma-glutamyltransferase, glucose, cholesterol, total protein, total bilirubin, blood urea nitrogen, creatine, and C-reactive protein by using a Piccolo point-of-care analyser and Biochemistry Panel Plus analyser discs (Abaxis).

Histopathology and immunohistochemistry. Necropsy was performed on all subjects. Tissue samples of all major organs were collected for histopathological and immunohistochemical examination, immersion-fixed in 10% neutral buffered formalin, and processed for histopathology as previously described[21]. For immunohistochemistry, specific anti-ZEBOV immunoreactivity was detected using an anti-

ZEBOV VP40 protein rabbit primary antibody (Integrated BioTherapeutics, Inc.) at a 1:4,000 dilution. In brief, tissue sections were processed for immunohistochemistry using the Dako Autostainer (Dako). Secondary antibody used was biotinylated goat anti-rabbit IgG (Vector Laboratories) at 1:200 followed by Dako LSAB2 streptavidin–horseradish peroxidase (Dako). Slides were developed with Dako DAB chromagen (Dako) and counterstained with haematoxylin. Non-immune rabbit IgG was used as a negative control.

19. Cooper, D. *et al.* Attenuation of recombinant vesicular stomatitis virus HIV-1 vaccine vectors by gene translocations and G gene truncation reduces neurovirulence and enhances immunogenicity in mice. *J. Virol.* **82,** 207–219 (2008). Medline CrossRef.
20. Witko, S. E. *et al.* An efficient helper-virus-free method for rescue of recombinant paramyxoviruses and rhadoviruses from a cell line suitable for vaccine development. *J. Virol. Methods* **135,** 91–101 (2006). Medline CrossRef.
21. Thi, E. P. *et al.* Marburg virus infection in nonhuman primates: Therapeutic treatment by lipid-encapsulated siRNA. *Sci. Transl. Med.* **6,** 250ra116 (2014).

a

b
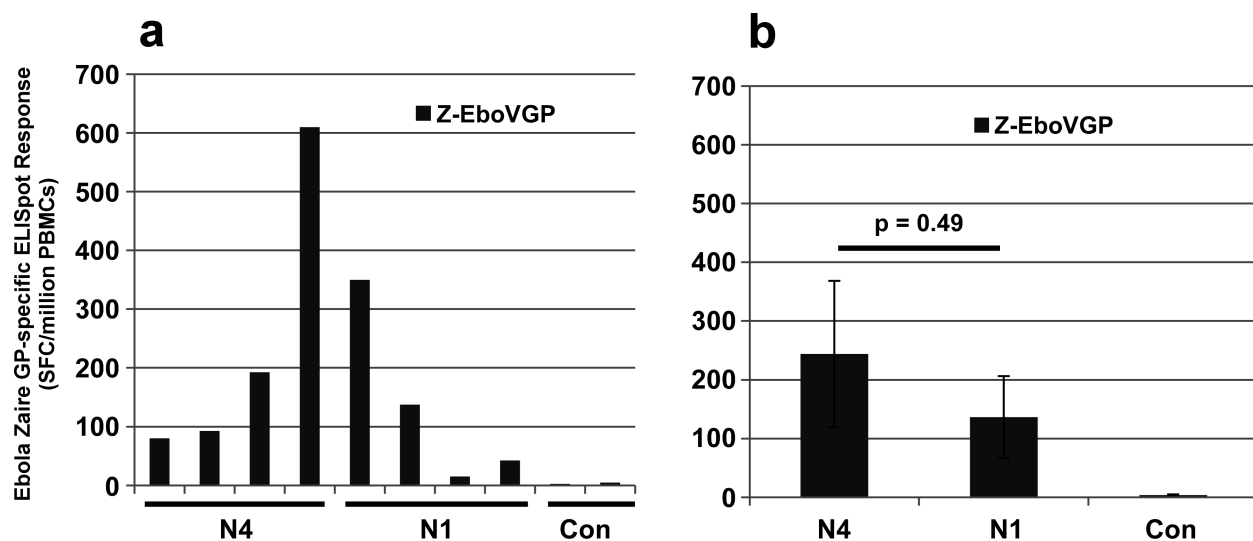
Extended Data Figure 1 | Relative immunogenicity of rVSV/ZEBOV vectors in cynomolgus macaques. At study day −28, cynomolgus macaques were immunized intramuscularly with 2×10^7 PFU of either N4 or N1 vectors. Ten days after a single immunization, PBMCs were prepared and ZEBOV GP-specific T-cell responses were quantified by IFN-γ ELISpot assay. **a**, ZEBOV GP-specific IFN-γ ELISpot responses in individual macaques. **b**, Average ZEBOV GP-specific IFN-γ ELISpot responses with s.e.m. indicated.

LETTER

Mutant MHC class II epitopes drive therapeutic immune responses to cancer

Sebastian Kreiter[1], Mathias Vormehr[2]*, Niels van de Roemer[2]*, Mustafa Diken[1], Martin Löwer[1], Jan Diekmann[1,3], Sebastian Boegel[1], Barbara Schrörs[1], Fulvia Vascotto[1], John C. Castle[1], Arbel D. Tadmor[1], Stephen P. Schoenberger[4], Christoph Huber[2], Özlem Türeci[1]§ & Ugur Sahin[1,2,3]§

Tumour-specific mutations are ideal targets for cancer immunotherapy as they lack expression in healthy tissues and can potentially be recognized as neo-antigens by the mature T-cell repertoire. Their systematic targeting by vaccine approaches, however, has been hampered by the fact that every patient's tumour possesses a unique set of mutations ('the mutanome') that must first be identified. Recently, we proposed a personalized immunotherapy approach to target the full spectrum of a patient's individual tumour-specific mutations[1]. Here we show in three independent murine tumour models that a considerable fraction of non-synonymous cancer mutations is immunogenic and that, unexpectedly, the majority of the immunogenic mutanome is recognized by CD4[+] T cells. Vaccination with such CD4[+] immunogenic mutations confers strong antitumour activity. Encouraged by these findings, we established a process by which mutations identified by exome sequencing could be selected as vaccine targets solely through bioinformatic prioritization on the basis of their expression levels and major histocompatibility complex (MHC) class II-binding capacity for rapid production as synthetic poly-neo-epitope messenger RNA vaccines. We show that vaccination with such polytope mRNA vaccines induces potent tumour control and complete rejection of established aggressively growing tumours in mice. Moreover, we demonstrate that CD4[+] T cell neo-epitope vaccination reshapes the tumour microenvironment and induces cytotoxic T lymphocyte responses against an independent immunodominant antigen in mice, indicating orchestration of antigen spread. Finally, we demonstrate an abundance of mutations predicted to bind to MHC class II in human cancers as well by employing the same predictive algorithm on corresponding human cancer types. Thus, the tailored immunotherapy approach introduced here may be regarded as a universally applicable blueprint for comprehensive exploitation of the substantial neo-epitope target repertoire of cancers, enabling the effective targeting of every patient's tumour with vaccines produced 'just in time'.

We recently reported comprehensive mapping of non-synonymous mutations of the B16F10 tumour by next-generation sequencing (Fig. 1a)[1]. Tumour-bearing C57BL/6 mice were immunized with synthetic 27mer peptides encoding the mutated epitope (mutation in position 14), resulting in T-cell responses which conferred *in vivo* tumour control. In continuation of that work, we now characterized the T-cell responses against the neo-epitopes, starting with those with a high likelihood of MHC I binding. Mice were vaccinated with synthetic 27mer peptides (Fig. 1b). Immunogenic mutations were identified by IFN-γ ELISpot of splenocytes and T-cell subtype was determined by cytokine release assay (Fig. 1a). About 30% of neo-epitopes were found to induce mutation-reactive cytokine-secreting

T cells. Surprisingly, responses against nearly all mutated epitopes (16/17, 95%) were CD4[+] (Fig. 1b, Extended Data Table 1).

To exclude bias associated with the peptide format, this experiment was repeated using *in vitro* transcribed (IVT) mRNA encoding the neo-epitopes. Also in this setting the majority of mutation-specific immune responses (10/12, ~80%) were conferred by CD4[+] T cells (Extended Data Fig. 1, Extended Data Table 1).

Recently, we identified over 1,680 non-synonymous mutations[2] in the colon carcinoma model CT26 in BALB/c mice. We selected 48 of these mutations in analogy to the B16F10 study based on good MHC class I binding ('low score' 0.1–2.1). The other half was deliberately chosen for poor binding ('high score' >3.9). In total, about 20% of mutated epitopes were found to be immunogenic in mice immunized with the respective RNA monotopes (Fig. 1c, Extended Data Table 2). In the 'low' MHC I score subgroup, but not in the 'high' score subgroup, several epitopes inducing CD8[+] T cells were identified (Fig. 1c right). MHC class II-restricted epitopes were represented in similar frequency in both subgroups, constituting the majority of CT26 immunogenic mutations (16/21, 80%).

Similarly, in the 4T1 mammary carcinoma model about 70% of the immunogenic epitopes determined by RNA monotope vaccines representing all 38 mutations of this model were recognized by CD4[+] T cells (Fig. 1d, Extended Data Table 3). In summary, we showed that in three independent mouse tumour models with different MHC backgrounds, a considerable fraction of non-synonymous cancer mutations is immunogenic and quite unexpectedly, the immunogenic mutanome is predominantly recognized by CD4[+] T cells.

To investigate whether MHC class II-restricted cancer mutations are good vaccine targets *in vivo*, we engineered pharmacologically optimized RNA (stabilizing elements in RNA sequence and liposomal formulation)[3–5] encoding B16-M30, one of the epitopes that elicited strong CD4[+] T-cell responses in the B16F10 tumour model. The mutated amino acid was essential for T-cell recognition, hence the wild type peptide was not recognized (Extended Data Fig. 2a). When B16F10 tumour-bearing C57BL/6 mice were repeatedly vaccinated with the B16-M30 RNA monotope, tumour growth was profoundly retarded (Fig. 2a). About two thirds of the neo-epitope RNA treated mice were still alive at day 100, while all the control RNA treated mice had died by 65 days. Depletion experiments in B16-M30 RNA treated mice revealed involvement of CD4[+] but not CD8[+] T-cells for therapeutic antitumour efficacy of the neo-epitope (Fig. 2a right).

Similarly, lung metastases of luciferase transduced B16F10 cells were efficiently eradicated with B16-M30 RNA but not control RNA in the vast majority of mice as shown by bioluminescence imaging (BLI) (Fig. 2b). Tumour infiltrating leukocytes purified from tumours of B16-M30 RNA immunized mice strongly reacted against B16-M30 (Fig. 2c). In tumours from neo-epitope-encoding RNA treated versus
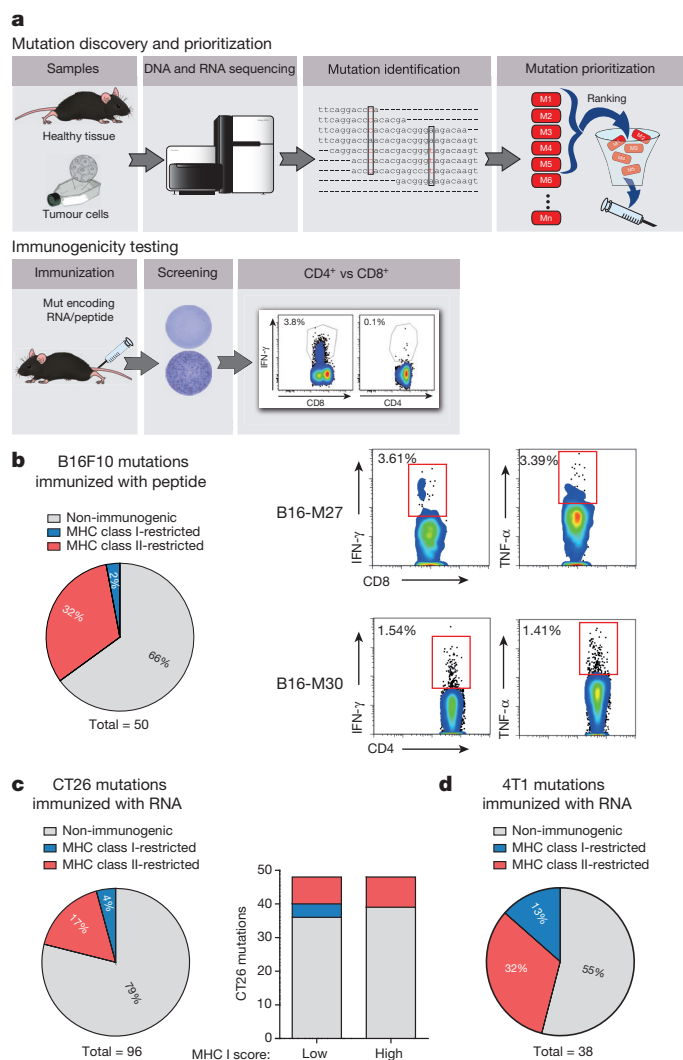
Figure 1 | Cancer-associated mutations are frequently immunogenic and pre-dominantly recognized by CD4+ T cells. a, Schematic describing mutation discovery and immunogenicity testing. **b–d**, Splenocytes of mice vaccinated with peptides and polyinosinic:polycytidylic acid (polyI:C) (**b**, B16F10, $n = 5$) or immunized with antigen-encoding RNA (**c**, CT26, $n = 5$; **d**, 4T1, $n = 3$) were tested for recognition of mutated peptides by ELISpot. Subsequent subtyping was performed via MHC II blockade or intracellular cytokine and CD4/CD8 surface staining. Pie charts represent the prevalence of non-immunogenic, MHC class I- or II-restricted mutated epitopes. **b**, Right, subtyping of mutation-specific T cells. **c**, Right, MHC restriction of neo-epitopes prioritized based on either good (0.1–2.1) or poor (>3.9) MHC I binding scores.

Figure 2 | Efficient tumour control and survival benefit in B16F10 melanoma with an RNA vaccine encoding a single mutated CD4+ T-cell epitope. a, Tumour growth (mean + s.e.m.) and survival (±CD4- or CD8-depleting antibodies) in untreated (control) or B16-M30 immunized C57BL/6 mice ($n = 10$) inoculated subcutaneously (s.c.) with B16F10. **b**, B6 albino mice ($n = 8$) developing lung metastases upon intravenous (i.v.) injection of B16F10-Luc were treated with B16-M30 or irrelevant RNA (control). Median tumour growth was determined by BLI as photons per second. **c**, Single-cell suspensions of B16F10 tumours of irrelevant (control) or B16-M30 RNA immunized mice ($n = 4$) were restimulated with B16-M30 or irrelevant peptide (vesicular stomatitis virus nucleoprotein, VSV-NP$_{52-59}$) and tested by ELISpot (mean + s.e.m.). Data pooled from two experiments. **d**, Frequency of infiltrating cells in s.c. B16F10 tumours ($n = 3$) left untreated (control) or vaccinated with B16-M30 RNA.

control mice, CD4+ as well as CD8+ T-cell infiltrates were significantly more abundant, whereas MDSCs and FoxP3+ T cells were significantly reduced (Fig. 2d). To benchmark antitumour efficacy, we tested immunogenic B16F10 neo-epitopes and a non-mutated TRP2-derived immunodominant epitope[6]. Whereas with TRP2 RNA, the two class I and three of the mutated class II neo-epitopes conferred a modest non-significant antitumour activity, three class II neo-epitopes mediated a significant inhibition of tumour growth (Extended Data Fig. 2d). Taken together, these data establish that a significant proportion of class II neo-epitopes have antitumour vaccine potency. Establishing B16-M30 as a novel rejection antigen in B16F10 our findings show that RNA encoding a single neo-epitope may give rise to functional T cells capable of targeting into the cancer lesion, controlling and even curing murine tumours. Our findings are in agreement with recent reports of the pivotal role of CD4+ T-cell immunity in the control of cancer[7,8].

The vast majority of mutations are unique to the individual patient. Hence, mutanome vaccines need to be individually tailored[9] and rapidly manufactured on-demand. This challenge can be viably addressed by RNA vaccine technology (Fig. 3a). At present, GMP-grade RNA is release-ready within 3 weeks. On another note, though we achieved tumour eradication in mice with a single mutation, to combine several mutations would be preferable to address tumour heterogeneity and immune editing, which mediate clinical failure of vaccines in humans[10,11]. We sought to integrate our insights into a concept which we call 'mutanome engineered RNA immunotherapy' (MERIT) (Fig. 3a).

To test this concept, we engineered RNA monotopes encoding four MHC class II (CT26-M03, CT26-M20, CT26-M27, CT26-M68) and one MHC class I (CT26-M19) restricted mutation from the CT26 model (see Extended Data Table 2) and a synthetic RNA pentatope encoding all five neo-epitopes connected by 10mer non-immunogenic glycine/serine linkers (Fig. 3a). In naive BALB/c mice the quantity of IFN-γ-producing T cells elicited by the pentatope was comparable (3 of 5) or even higher than that evoked by the respective monotope (Extended Data Fig. 3a).

In BALB/c mice with CT26 luciferase-transfected (CT26-Luc) lung metastases vaccinated repeatedly with a mixture of two RNA
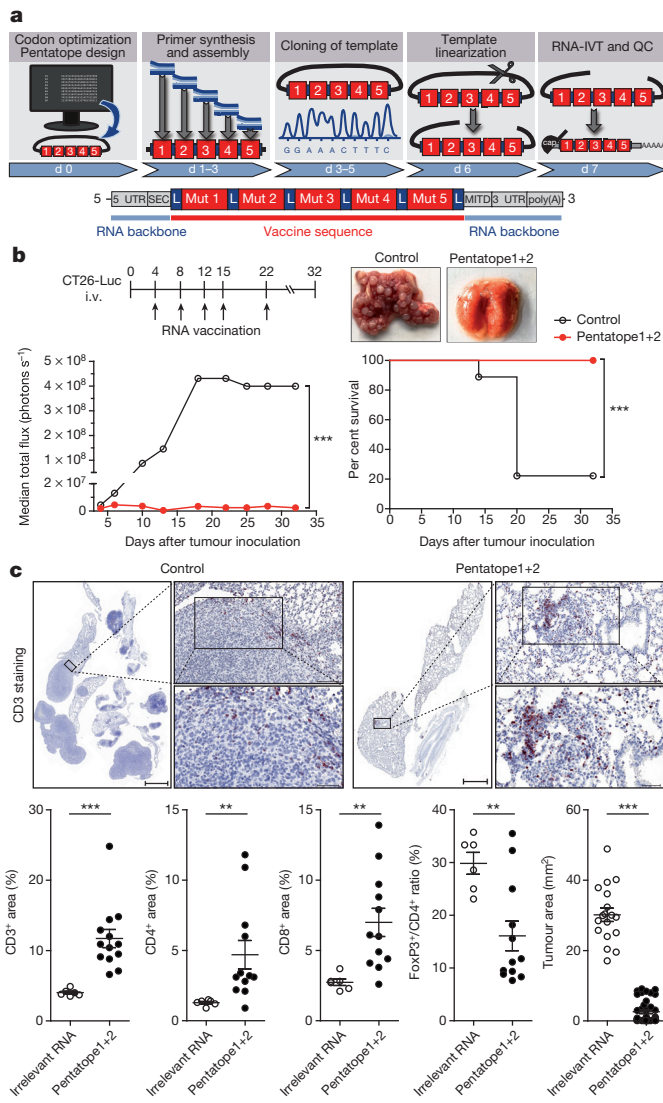
Figure 3 | RNA pentatope immunization confers disease control and survival benefit in murine tumours. **a**, Engineering of a poly-neo-epitope RNA. **b**, BALB/c mice ($n = 10$) developing lung metastases upon i.v. injection of CT26-Luc were treated with a mixture of two pentatopes or left untreated (control). The median tumour growth by BLI, survival data and lungs from treated animals are shown. **c**, Upper panel, CD3 stained lung tissue sections. Scale bars: 1,000 μm (scan), 100 μm (top), 50 μm (bottom). Lower panel, proportional lymphocyte areas in lung tumour tissue of control ($n = 6$) or pentatope-treated (CD3: $n = 14$; CD4, CD8, FoxP3: $n = 12$) animals. Lower panel right, tumour area (mean ± s.e.m.) in sections of control ($n = 18$) and pentatope1+2-treated ($n = 39$) mice.

pentatopes (3 MHC class I- and 7 class II-restricted epitopes, Extended Data Table 4) including the mutations tested in the previous experiment, tumour growth was significantly inhibited (Fig. 3b). At day 32 all mice in the RNA pentatope group were alive whereas 80% of the control mice had already died. Dissection of the antitumour activity of single RNA pentatopes in the CT26 model revealed that RNA pentatope 2 has a very strong antitumour activity, whereas pentatope 1 is modestly active (Extended Data Fig. 3b). A subsequent study confirmed the antitumour activity of pentatope 2, and showed its significant decrease upon CD40L blockade and complete loss by CD8$^+$ T-cell depletion (Extended Data Fig. 3c). Tumour load in the vaccinated as compared to untreated mice was significantly lower as shown by post mortem macroscopic (Fig. 3b), histological (Fig. 3c upper panel) and computerized image analysis (Fig. 3c lower panel right) of tissue sections. CD3$^+$ T-cell infiltrates in tumour lesions of

pentatope RNA vaccinated mice, in contrast to findings in controls, were significantly brisker as compared to the surrounding lung tissues (Fig. 3c lower panel left). Moreover, in independent experiments we found a significant increase of CD4$^+$ and CD8$^+$ T cells in tumour lesions of RNA pentatope immunized mice but a lower FoxP3/CD4 ratio as compared to tumours of mice treated with irrelevant RNA (Fig. 3c lower panel, Extended Data Fig. 3d).

Taken together our data indicate that with a poly-neo-epitope encoding RNA vaccine T cells against each single epitope are elicited. These T cells target tumour lesions, recognize their mutated targets, reshape the cellular composition of the tumour microenvironment and result in efficient tumour control *in vivo*.

The current paradigm for selecting mutations for immunization is to employ MHC class I binding scores for enrichment of mutated epitope candidates[12–14] which can elicit CD8$^+$ responses and tumour rejection. Our findings indicate that MHC class II presented neo-epitopes may be of higher interest for a MERIT approach. In fact, a correlation analysis revealed that immunogenic mutations have a significantly better MHC class II binding score as compared to non-immunogenic ones (Fig. 4a). As most cancers lack MHC class II expression, effective recognition of neo-epitopes by CD4$^+$ T cells should depend on presentation of released tumour antigens by antigen presenting cells (APCs). This is most efficient for highly expressed antigens[15]. Thus, we implemented an algorithm combining good MHC class II binding with abundant expression of the mRNA encoding the neo-epitope based on confirmed mutated RNA sequencing reads normalized to the overall read count (NVRC: normalized variant read counts). To test the impact of predicted MHC class II binding affinity, we ranked CT26 mutanome data with this algorithm and selected the top ten mutations ('ME' mutations in Extended Data Table 5) predicted to be the best MHC class II binders among the most abundant candidate epitopes (NVRC ≥ 60). The control comprised ten mutations with abundant expression only ('E' mutations in Extended Data Table 5). These neo-epitopes were used without any further pre-validation or immunogenicity testing to engineer two RNA pentatopes for each group (P$_{ME}$ and P$_E$ pentatopes). In mice with established CT26-Luc lung tumours, P$_{ME}$ induced a much stronger T-cell response as compared to P$_E$ pentatopes (Fig. 4c). Analysis of immune responses proved the presence of multiple immunogenic MHC class II neo-epitopes in the P$_{ME}$ RNA pentatopes (Extended Data Fig. 4). Established lung metastases were completely rejected in almost all P$_{ME}$ mice whereas P$_E$ pentatopes were not able to control tumour growth (Fig. 4b). An independent study confirmed the strong antitumour activity of the pentatope P$_{ME}$ and showed loss of the antitumour effect upon anti-CD40L and anti-CD8 treatments (Fig. 4d).

Antigen-specific T$_H$ cells promote the cross-priming of tumour-specific cytotoxic T lymphocyte (CTL) responses by CD40 ligand-mediated activation of dendritic cells. If T$_H$ cells recognize their antigen on the same APC (cross-)presenting an CTL epitope, a diversified CTL response may result[16]. In support of this notion, we detected strong CD8$^+$ T-cell responses against gp70-AH1, a well characterized immunodominant CTL epitope of the endogenous murine leukaemia virus-related cell surface antigen in the blood and spleen of mice immunized with P$_{ME}$ but not P$_E$ pentatopes (Fig. 4e). This indicates that cancer neo-epitope specific T$_H$ cells, in analogy to viral neo-antigen specific T cells, may exert their antitumour function by augmentation of CTL responses through epitope spreading.

In summary, we show that MHC class II-restricted T-cell epitopes are more abundant than previously appreciated in the cancer mutanome and can be targeted by customized RNA-based poly-neo-epitope vaccines with substantial therapeutic effect in mouse tumour models. A recent study reports that about 0.5% of mutations induce spontaneous CD4$^+$ T-cell immune responses in human tumours[17]. Even by the most conservative estimate, our findings indicate that the proportion of relevant neo-epitope vaccine targets recognized by CD4$^+$ T-cells is a full log higher than that reported to induce
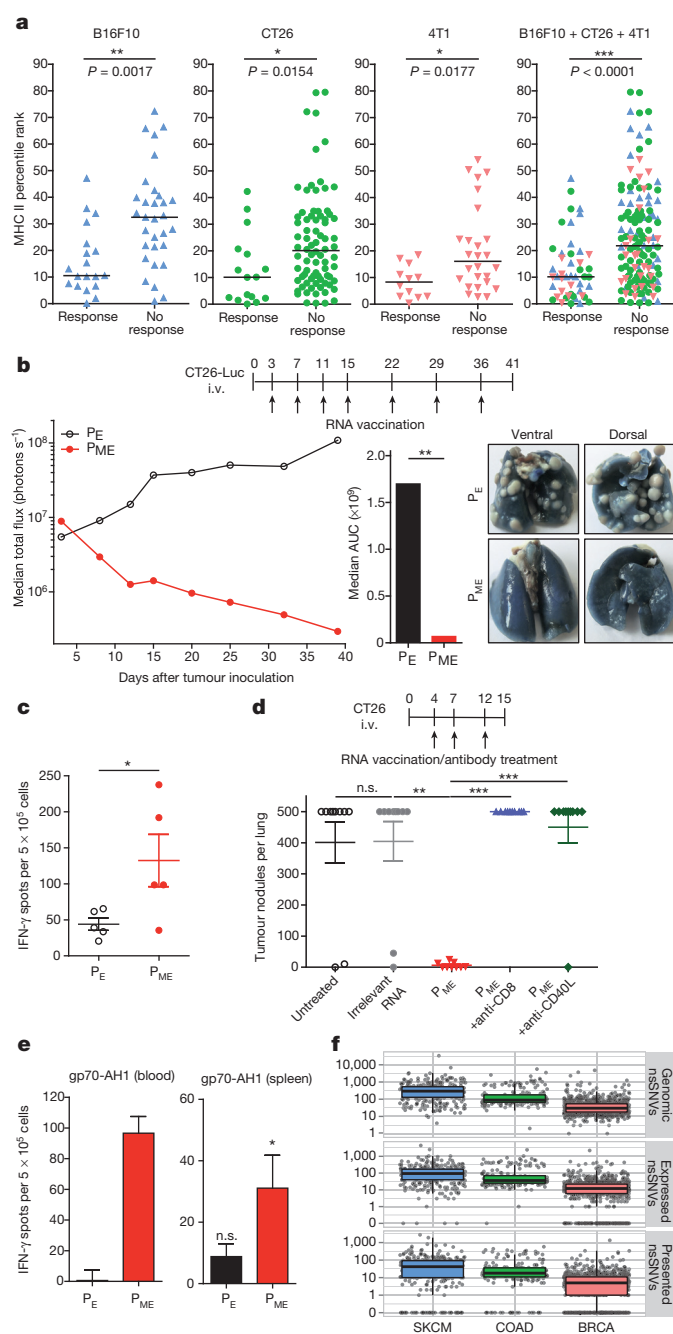
Figure 4 | RNA pentatope vaccines with mutations selected for *in silico* predicted favourable MHC class II binding and abundant expression confer potent antitumour control. a, Comparison of median MHC II binding scores of immunogenic (Response) and non-immunogenic (No response) mutated 27mers. **b**, Highly expressed mutations were selected with ('ME') or without ('E') considering MHC class II binding score. Ten mutations (two pentatopes) per category were used for vaccination of CT26-Luc tumour-bearing mice ($n = 10$). Tumour growth, area under the curve (AUC) at day 40 and ink-treated lungs are shown. **c**, Mice ($n = 5$) were analysed for T-cell responses against the RNA pentatopes via ELISpot (mean ± s.e.m. subtracted by an irrelevant RNA control). **d**, CT26 tumour nodules per lung of untreated mice or mice ($n = 10$) injected with irrelevant or P_{ME} (±CD8 depletion or CD40L blocking) RNA. **e**, T-cell responses against gp70$_{423-431}$ (gp70-AH1) determined via ELISpot in blood (pooled from 5 mice at day 20) and spleen ($n = 5$). Background (medium control) subtracted mean ± s.e.m. shown. **f**, Genomic, expressed and predictively presented (HLA-DRB1, IEDB rank <10) non-synonymous single nucleotide variations (nsSNVs) derived from human cancer samples (TCGA). SKCM, skin cutaneous melanoma; COAD, colon adenocarcinoma; BRCA, breast invasive carcinoma.

spontaneous CD4[+] T-cell responses. The reason might be that spontaneous immunogenicity is not driven by antigenicity alone, but is a function of many factors including tumour biology and local immunosuppression. Both B16 (Extended Data Fig. 2e) and CT26 are tumours with strong immunosuppressive properties, such that they are not able to induce protective immunity unless transfected with immunomodulators such as GM-CSF. As our study shows that vaccination with CD4[+] T-cell neo-epitopes identified by exome sequencing counteracts the immunosuppressive tumour microenvironment in these models resulting in rejection of established tumours, one key prediction from our work is that such 'poorly immunogenic' tumours can be successfully targeted through induction of mutation-specific CD4[+] T-cell responses.

A simple explanation for this remarkable frequency of CD4[+] T-cell recognition of mutations may be less stringent length and sequence requirement for peptides binding to MHC class II molecules as compared to MHC class I epitopes increasing the likelihood that a given mutation is found within a presented peptide[18].

The first evidence of spontaneous CD8[+] and CD4[+] T-cell responses directed against mutated gene-products in cancer patients was generated in the 1990s[19,20]. Recent publications have renewed enthusiasm for the enormous potential of mutation-specific T cells to confer antitumour activity in cancer patients[7,8,21]. As elegantly proven in a murine carcinogen-induced sarcoma model[22] there is evidence for a direct link between T-cell recognition of mutant neo-epitopes and clinical response to checkpoint blockade treatment[14,23].

To assess whether the principles unraveled in the mouse models for melanoma, colon and breast cancer are true in the human setting, we analysed mutation and RNA-Seq data in the same human cancer types provided by The Cancer Genome Atlas (TCGA). For all three human cancers we confirmed the abundance of mutations predicted to bind to MHC class II (Fig. 4f).

The MERIT approach we presented here by integrating advances in the field of next-generation sequencing, computational immunology and synthetic genomics provides the technology for comprehensive exploitation of the neo-epitope target repertoire. The approach might be particularly useful to reshape the tumour microenvironment in patients who lack T-cell infiltration in their tumours and applicable as a standalone or combination therapy to increase the clinical success rate of checkpoint blockade therapy that depends on pre-existing immunity[24–26]. Targeting multiple mutations at once may in theory pave the way to solve critical problems in current cancer drug development such as clonal heterogeneity and antigen escape[10,12].

Based on the findings of this study and of our prior work, a first-in-concept trial in melanoma patients[1,27,28] has been initiated and is recruiting (NCT02035956), confirming that 'just in time' production of a poly-neo-epitope mRNA cancer vaccine is in fact feasible.

1. Castle, J. C. *et al.* Exploiting the mutanome for tumor vaccination. *Cancer Res.* **72**, 1081–1091 (2012).
2. Castle, J. C. *et al.* Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* **15**, 190 (2014).
3. Holtkamp, S. *et al.* Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* **108**, 4009–4017 (2006).
4. Kreiter, S. *et al.* Increased antigen presentation efficiency by coupling antigens to MHC class I trafficking signals. *J. Immunol.* **180**, 309–318 (2008).
5. Kuhn, A. N. *et al.* Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses *in vivo*. *Gene Ther.* **17**, 961–971 (2010).
6. Bloom, M. B. *et al.* Identification of tyrosinase-related protein 2 as a tumor rejection antigen for the B16 melanoma. *J. Exp. Med.* **185**, 453–459 (1997).

7. Schumacher, T. *et al.* A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512**, 324–327 (2014).
8. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).
9. Britten, C. M. *et al.* The regulatory landscape for actively personalized cancer immunotherapies. *Nature Biotechnol.* **31**, 880–882 (2013).
10. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
11. Koebel, C. M. *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903–907 (2007).
12. Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* **482**, 400–404 (2012).
13. Robbins, P. F. *et al.* Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nature Med.* **19**, 747–752 (2013).
14. van Rooij, N. *et al.* Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J. Clin. Oncol.* **31**, e439–e442 (2013).
15. Shen, Z. *et al.* Cloned dendritic cells can present exogenous antigens on both MHC class I and class II molecules. *J. Immunol.* **158**, 2723–2720 (1997).
16. Schoenberger, S. P. *et al.* T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions. *Nature* **393**, 480–483 (1998).
17. Linnemann, C. *et al.* High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4$^+$ T cells in human melanoma. *Nature Med.* **21**, 81–85 (2015).
18. Arnold, P. Y. *et al.* The majority of immunogenic epitopes generate CD4$^+$ T cells that are dependent on MHC class II-bound peptide-flanking residues. *J. Immunol.* **169**, 739–749 (2002).
19. Wolfel, T. *et al.* A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* **269**, 1281–1284 (1995).
20. Wang, R. F. *et al.* Cloning genes encoding MHC class II-restricted antigens: mutated CDC27 as a tumor antigen. *Science* **284**, 1351–1354 (1999).
21. Lu, Y. C. *et al.* Mutated PPP1R3B is recognized by T cells used to treat a melanoma patient who experienced a durable complete tumor regression. *J. Immunol.* **190**, 6034–6042 (2013).
22. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
23. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
24. Herbst, R. S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).
25. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
26. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
27. Castle, J. C. *et al.* Mutated tumor alleles are expressed according to their DNA frequency. *Sci. Rep.* **4**, 4743 (2014).
28. Löwer, M. *et al.* Confidence-based somatic mutation evaluation and prioritization. *PLOS Comput. Biol.* **8**, e1002714 (2012).

METHODS

Cell lines and mice. Female 8–12 weeks old C57BL/6, BALB/c mice (Janvier Labs) and C57BL/6BrdCrHsd-Tyr[c] mice (B6 albino, Harlan) were kept in accordance with federal policies on animal research at the University of Mainz. B16F10 melanoma cell line, CT26 colon carcinoma cell line and 4T1-luc2-tdtomato (4T1-Luc) cells were purchased in 2010, 2011 and 2011, respectively (ATCC CRL-6475 lot no. 58078645, ATCC CRL-2638 lot no. 58494154, Caliper 125669 lot no. 101648). Firefly-luciferase-expressing CT26-Luc and B16F10-Luc cells were lentivirally transduced. Master and working cell banks were generated immediately upon receipt, of which third and fourth passages were used for tumour experiments. Cells were tested for mycoplasma every 3 months. Reauthentification of cells was not performed since receipt.

Next-generation sequencing and data processing. As described previously[1], exome capture from mouse tumour cells and tail tissue samples of BALB/c or C57BL/6 mice were sequenced in triplicate (except 4T1-Luc in duplicate). Oligo (dT)-based RNA sequencing libraries for gene expression profiling were prepared in triplicate. Libraries were sequenced on an Illumina HiSeq2000 to generate 50 nucleotide single-end (B16F10) or 100 nucleotide paired-end (CT26, 4T1-Luc) reads, respectively. Gene expression values were determined by counting reads overlapping transcript exons and junctions, and normalizing to RPKM expression units (reads which map per kilobase of transcript length per million mapped reads). Mutation expression was determined by normalization of mutated RNA reads to the total mapped read counts multiplied by 100 million (normalized variant read counts, NVRC).

Mutation selection, validation and prioritization. As described previously[1,2,28], mutations were selected based on following criteria: (1) present in the respective tumour cell line sequencing triplicates and absent in the corresponding healthy tissue sample triplicates, (2) occur in a RefSeq transcript, (3) cause non-synonymous changes, and (4) occurrence in expressed genes of tumour cell lines (median RPKM across replicates). For validation, mutations were amplified from DNA from the respective cell lines or mice tail tissue and subjected to Sanger sequencing. DNA-derived mutations were classified as validated if confirmed by either Sanger sequencing or the RNA-Seq reads. No confirmation via Sanger sequencing and immunogenicity testing was performed for experiments in Fig. 4. For experiments shown in Fig. 1 mutated epitopes were prioritized according to their MHC class I binding predicted by the consensus method (version 2.5) of the Immune Epitope Database (http://www.iedb.org). Mutations shown in Fig. 4b–e were selected based on either their expression (NVRC) alone or together with their predicted MHC class II peptide binding capability (IEDB consensus method version 2.5). Retrospective analysis of MHC II binding prediction shown in Fig. 4a was determined with IEDB consensus method version 2.12. For analysis of mutations in human tumours, DNA sequencing data of skin cutaneous melanoma (SKCM, $n = 308$), colon adenocarcinoma (COAD, $n = 192$) or breast invasive carcinoma (BRCA, $n = 872$) retrieved from The Cancer Genome Atlas (TCGA) (August 2014) was filtered to obtain genomic non-synonymous point mutations (nsSNVs). RNA-Seq data (TCGA) of tumour samples with identified genomic mutations was used to define expressed nsSNVs (STAR aligner, reference genome: GRCh37/hg19, max. mismatch ratio of 2%). The reads mapping to the reference genome were intersected with the UCSC gene model database. To predict MHC II binding of expressed neo-epitopes, seq2HLA[29] was employed to identify the patients' 4-digit HLA class II (HLA-DQA1, HLA-DQB1, HLA-DRB1) type. The IEDB consensus binding prediction (version 2.12) was used to predict MHC class II binding from a 27mer peptide and the patients HLA-DRB1 alleles. As recommended from IEDB, neo-epitopes with a percentile rank below 10% were considered as binders.

Synthetic RNA and synthetic peptides. Non-synonymous mutations were studied in the context of the respective 27mer amino acid epitope with the mutated amino acid in the centre (position 14). Mutated peptides were synthesized together with gp70-AH1 (gp70$_{423–431}$) and control peptides (vesicular stomatitis virus nucleoprotein (VSV-NP$_{52–59}$), tyrosinase-related protein 2 (TRP2$_{180–188}$) by JPT Peptide Technologies GmbH. Alternatively, sequences encoding mutated 27mer peptides were cloned into the pST1-Sp-MITD-2hBgUTR-A120 backbone[3] featuring sequence elements for pharmacologically optimized synthetic RNA in terms of translation efficiency and MHC class I/II processing of epitopes either as monotopes or fused to each other by sequences encoding 10 amino acid long glycine-serine linker in between resulting in pentatopes. Linearization of these plasmid constructs, in vitro translation (IVT) and purification are described in detail elsewhere[3].

Mouse models. For immunogenicity studies of mutated epitopes age-matched female C57BL/6 or BALB/c mice were vaccinated on day 0, 3, 7 and 14 (RNA immunization) or day 0 and 7 (peptide immunization). The read out was performed five to six days after the last immunization (see also Enzyme-linked ImmunoSpot and Flow cytometric analysis method sections). Vaccination was performed either by i.v. injection of 200 μl (20 μg per mutation for B16F10, 40 μg per mutation for CT26) RNA complexed with cationic lipids (manuscript in preparation) or subcutaneous (s.c.) injection of 100 μg synthetic peptide and 50 μg poly(I:C) formulated in PBS (200 μl total volume) into the lateral flank. Two mutations per mouse were tested ($n = 5$ for B16F10, $n = 3$ for CT26). For confirmation of immunogenic mutations and subtyping, mice were vaccinated against a single mutation ($n = 5$).

For therapeutic tumour experiments C57BL/6 mice were inoculated s.c. with 1×10^5 B16F10 melanoma cells into the flank and randomly distributed into treatment groups. Tumour volume was measured unblinded with a caliper and calculated using the formula $(A \times B^2)/2$ (A as the largest and B the smallest diameter of the tumour). Tumour growth was documented as mean tumour size with standard error disregarding single distant outliers.

In lung metastasis experiments 5×10^5 CT26-Luc or 2×10^5 CT26 cells were injected into the tail vein of BALB/c mice or 1.5×10^5 B16F10-Luc tumour cells into B6 albino mice. Tumour growth of luciferase transgenic cells was traced unblinded by bioluminescence imaging after i.p. injection of an aqueous solution of D-luciferin (250 μl, 1.6 mg, BD Bioscience) on an IVIS Lumina (Caliper Life Sciences). Five minutes after injection emitted photons were quantified. In vivo bioluminescence in regions of interest (ROI) were quantified as total flux (photons s^{-1}) (IVIS Living Image 4.0). Mice were randomized based on their total flux values (ANOVA-P method, Daniel's XL Toolbox V6.53). CT26 lung tumour burden was quantified unblinded after tracheal ink (1:10 diluted in PBS) injection and fixation with Fekete's solution (5 ml 70% ethanol, 0.5 ml formalin, and 0.25 ml glacial acetic acid). In therapeutic experiments mice were administered repeated doses of either monotope (40 μg), pentatope RNA (in total 40 μg) or equimolar amounts of irrelevant RNA. In some experiments repeated doses (200 μg per mouse i.p.) of CD8-depleting (clone YTS191, BioXcell), CD4-depleting (clone YTS169.1, BioXcell) or CD40L-blocking (clone MR1) antibodies were administered. The experimental group sizes were approved by the regulatory authorities for animal welfare after being defined to balance statistical power, feasibility and ethical aspects.

Enzyme-linked ImmunoSpot (ELISpot). As previously described[30], 5×10^5 splenocytes were cultured over night at 37 °C in anti-INF-γ (10 μg ml^{-1}, clone AN18, Mabtech) coated Multiscreen 96-well plates (Millipore) and cytokine secretion was detected with an anti-IFN-γ antibody (1 μg ml^{-1}, clone R4-6A2, Mabtech). For stimulation either 2 μg ml^{-1} peptide was added or spleen cells were coincubated with 5×10^4 syngeneic bone-marrow-derived dendritic cells (BMDC) transfected with RNA. For analysis of tumour infiltrating lymphocytes, single-cell suspensions of lung metastasis were rested overnight to get rid of living tumour cells via plastic adherence. Viable cells were separated via density gradient centrifugation and added to the ELISpot plate. For analysis of T-cell responses in peripheral blood, PBMC were isolated via density gradient centrifugation, counted and restimulated by addition of peptide and syngeneic BMDC. Subtyping of T-cell responses was performed with an MHC class II blocking antibody (20 μg ml^{-1}, clone M5/114, BioXcell). All samples were tested in duplicates or triplicates.

Flow cytometric analysis. In the presence of Brefeldin A (Sigma-Aldrich) 2×10^6 splenocytes were stimulated with 2×10^5 RNA-transfected BMDC or 2 μg ml^{-1} peptide. Splenocytes treated with phorbol 12-myristate 13-acetate (PMA, 0.5 μg ml^{-1}, Sigma-Aldrich) and Ionomycin (1 μg ml^{-1}, Sigma-Aldrich) served as a positive control. Cells were incubated for 5 h at 37 °C, stained for CD4$^+$ and CD8$^+$ cell surface markers, permeabilized and fixed using BD Cytofix/Cytoperm according to the manufacturer's protocol. Thereafter cells were stained for INF-γ, TNF-α and IL-2 cytokines (BD Biosciences). Cytokine secretion among CD4$^+$ or CD8$^+$ T cells in stimulated samples was compared to control samples (medium, irrelevant RNA or irrelevant peptide) in order to determine the responding T-cell subtype ($n = 5$). Tumour infiltrating leucocytes were prepared from subcutaneous B16F10 tumours fifteen to twenty days after inoculation. Tumours were harvested and minced into pieces of 1–2 mm diameter. The resulting cell suspension was harvested, filtered through a 70-μm cell strainer, washed two times and stained for CD4, CD8, Gr-1 and CD11b surface marker. Intracellular FoxP3 staining was performed according to the manufacturer's protocol (Mouse Foxp3 Buffer Set, BD). Samples were acquired on a BD FACSCanto II.

Immune histochemistry. Lungs of CT26 tumour bearing mice were fixed overnight in 4% phosphate buffered formaldehyde solution (Carl Roth) and embedded in paraffin. 50-μm consecutive sections (3 per mouse) were stained for CD3 (clone SP7, Abcam), CD4 (clone 1, catalogue no. 50134-M08H, Sino Biologinal) and FoxP3 (polyclonal, catalogue no. NB100-39002, Novus Biologicals) following detection by a HRP-conjugated antibody (Poly-HRP-anti-rabbit IgG, ImmunoLogic) and the corresponding peroxidase substrate (Vector Nova Red, Vector Laboratories) and counterstained with hematoxylin. CD3$^+$, CD4$^+$, FoxP3$^+$ and tumour areas were captured on an Axio Scan.Z1 (Zeiss) and manually pre-defined tumour and lung regions were quantified via computerized image analysis software (Tissue Studio 3.6.1, Definiens). CD8$^+$ area was calculated by subtracting CD4

stained area from CD3$^+$ area. For comparison of tumour areas between control and pentatope1+2-treated animals, tumour-free sections were excluded.

Immunofluorescence staining. 8-μm sections of cryo-conserved organs were attached on Superfrost slides, dried overnight at room temperature and fixed in 4% para-formaldehyde (PFA) for 10 min at room temperature in the dark. Sections were washed 3 times with PBS and blocked using PBS supplemented with 1% BSA, 5% mouse serum, 5% rat serum and 0.02% Nonident for 1 h at room temperature in the dark. Fluorescent labelled antibodies (FoxP3, clone FJK-16 s, eBioscience; CD8, clone 53-6.7, BD; CD4, clone RM4-5, BD) were diluted in staining buffer (PBS supplemented with 1% BSA, 5% mouse serum and 0.02% Nonident) and sections were stained overnight at 4 °C. After washing twice with washing buffer (PBS supplemented with 1% BSA and 0. 02% Nonident) and once with PBS, slides were stained for 3 min with Hoechst (Sigma), washed 3 times with PBS, once with distilled water and mounted using Mounting Medium Flouromount G (eBioscience). Immunofluorescence images were acquired using an epifluorescence microscope (ApoTome, Zeiss). Tumour, CD4, CD8 and FoxP3 stained areas were quantified within manually pre-defined tumour regions via computerized image analysis software (Tissue Studio 3.6.1., Definiens). The proportion of marker positive cells in comparison to DAPI positive cells was calculated.

Statistics. Means were compared by using Student's t-test for hypothesis testing to compare individual treatment and corresponding control groups. In case of significantly different variances (F-test, alpha = 0.05) Welch's correction was used. Mann–Whitney U test was applied if data sets failed the Pearson omnibus normality test (alpha = 0.05). Tumour growth was compared by calculating the area under the tumour growth curve (AUC) for single mice. Statistical differences in medians between two groups were calculated with a nonparametric Mann–Whitney U test. Survival benefit was determined with the log-rank test. All analyses were two-tailed (except Fig. 4c, e) and carried out using GraphPad Prism 5.03. n.s.: $P > 0.05$, *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$. Grubb's test was used for identification of outliers (alpha = 0.05). No statistical methods were used to pre-determine sample size.

29. Boegel, S. et al. A catalog of HLA type, HLA expression, and neo-epitope candidates in human cancer cell lines. *OncoImmunology* **3,** e954893 (2014).
30. Kreiter, S. et al. Intranodal vaccination with naked antigen-encoding RNA elicits potent prophylactic and therapeutic antitumoral immunity. *Cancer Res.* **70,** 9031–9040 (2010).
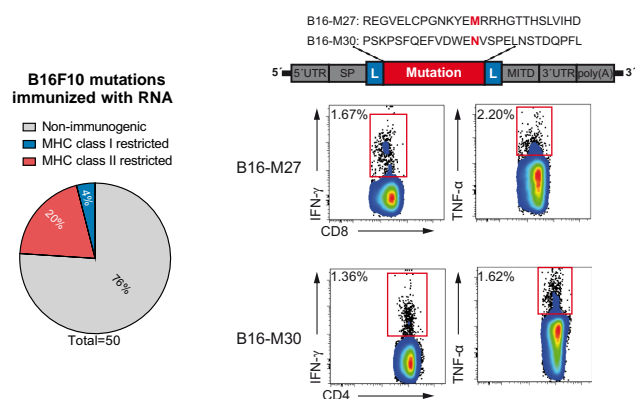
Extended Data Figure 1 | Non synonymous cancer-associated mutations are frequently immunogenic and pre-dominantly recognized by CD4[+] T cells. T-cell responses obtained by vaccinating C57BL/6 mice with antigen-encoding RNA in the B16F10 tumour model ($n = 5$). Left, prevalence of non-immunogenic, MHC-class-I- or class-II-restricted mutated epitopes. Right, detection and typing of mutation-specific T cells (individual epitopes shown in Extended Data Table 1).
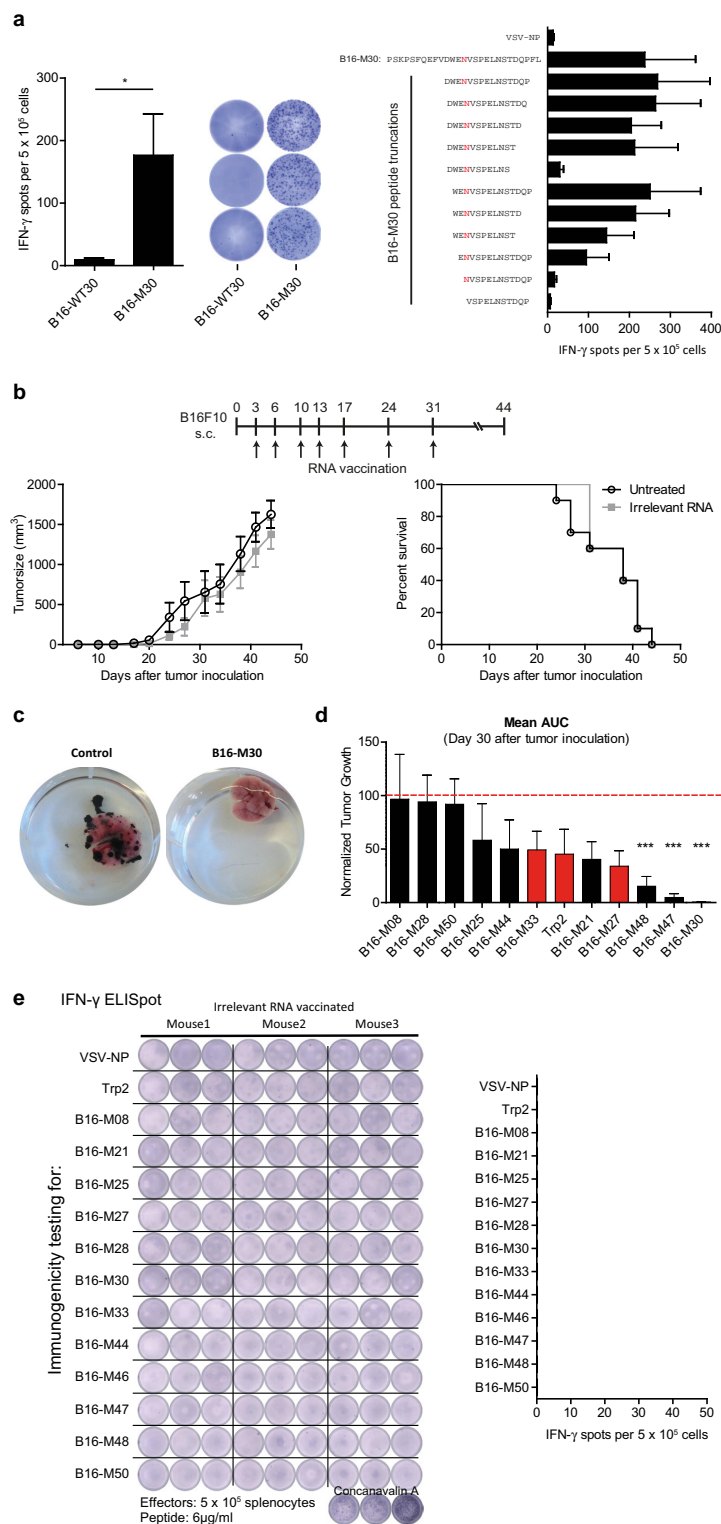
Extended Data Figure 2 | Mutant epitope-specific T cells induced by RNA vaccination control tumour growth. **a**, Splenocytes of mice ($n = 5$) vaccinated with B16-M30 RNA were tested by ELISpot for recognition of mutated peptides as compared to the corresponding wild-type (B16-WT30) sequence. Right, testing of truncated variants of B16-M30 (mean + s.e.m.). **b**, Mean ± s.e.m. tumour growth (left) and survival (right) of C57BL/6 mice ($n = 10$) inoculated s.c. with B16F10 and left untreated (control) or injected i.v. with irrelevant RNA. **c**, Lungs of B16F10-Luc tumour bearing mice shown in Fig. 2b (day 27 after tumour inoculation). **d**, Therapeutic antitumour activity against B16F10 tumours in mice (B16-M27, Trp2 $n = 8$; B16-M30 $n = 7$; others $n = 10$) conferred by immunization with epitopes encoding immunogenic B16F10 mutations or an immunodominant wild type Trp2 epitope[6]. The area under the tumour growth curve at day 30 after tumour inoculation was normalized to untreated control mice and depicted as mean ± s.e.m. Red and black columns represent mutations recognized by CD8[+] or CD4[+] T cells, respectively. **e**, Spontaneous immune responses in splenocytes of irrelevant RNA treated B16F10 tumour bearing C57BL/6 mice ($n = 3$) were tested by ELISpot for recognition of peptides (mean + s.e.m.).
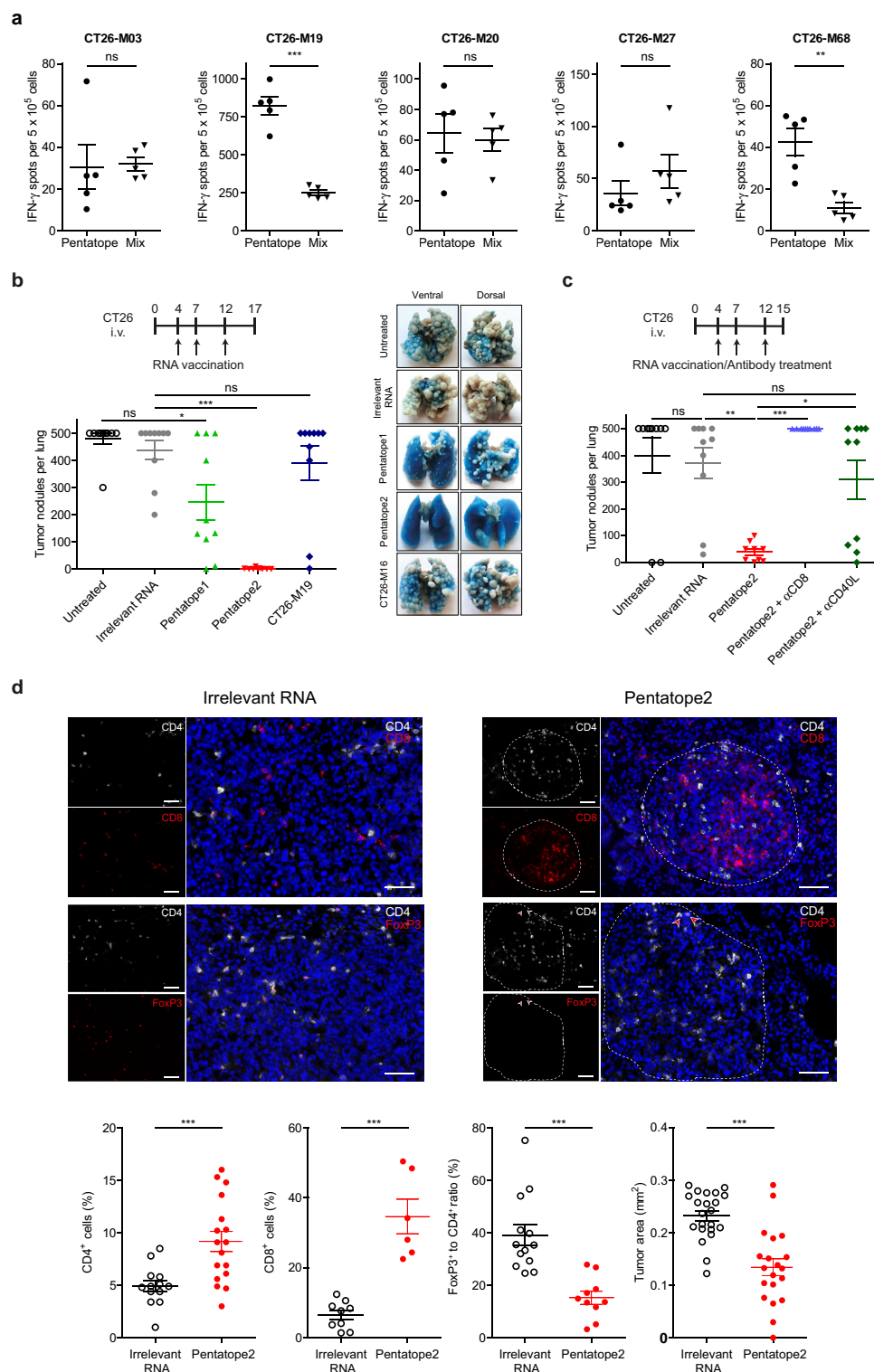
Extended Data Figure 3 | Mechanism of antitumour activity of mutation specific poly-epitope vaccines in CT26 tumour-bearing mice. a, BALB/c mice ($n = 5$) were vaccinated either with pentatope (35 μg) or the corresponding mixture of five RNA monotopes (7 μg each). T-cell responses in peptide-stimulated splenocytes of mice were measured *ex vivo* via ELISpot (medium control subtracted mean ± s.e.m.). **b, c**, BALB/c mice ($n = 10$) were inoculated i.v. with CT26 tumour cells and left untreated or injected with irrelevant, CT26-M19 or pentatope1 or 2 RNA in absence (**b**) or presence of a CD8 T cell depleting antibody or a CD40L blocking antibody (**c**). Mean ± s.e.m. of tumour nodules per lung are shown. **d**, Immunofluorescence analyses of tumour-infiltrating lymphocytes in pentatope2-vaccinated mice. Upper panel, lung tumour tissue stained for CD4 and CD8 or CD4 and FoxP3. Scale bar, 50 μm. Lower panel left, proportion of infiltrating cells in sections of irrelevant (CD4: $n = 13$; CD8 = 9; FoxP3: $n = 13$) or pentatope (CD4: $n = 17$; CD8: $n = 6$; FoxP3: $n = 10$) RNA-treated animals. Lower panel right, tumour area in sections of control ($n = 22$) and pentatope2-treated ($n = 20$) animals (mean ± s.e.m.).
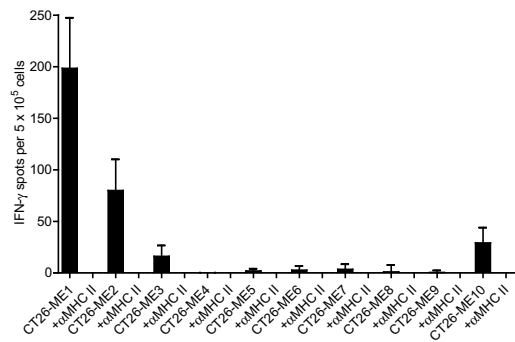
Extended Data Figure 4 | Immunogenicity testing of P_{ME} pentatope-encoded mutations. Splenocytes of P_{ME} RNA vaccinated BALB/c mice ($n = 6$) were tested *ex vivo* for recognition of peptides representing the mutated 27mer sequences represented in P_{ME} pentatopes with or without addition of an MHC class II-blocking antibody. Mean + s.e.m. of background (medium control) subtracted responses are shown.

Extended Data Table 1 | Immunogenic B16F10 mutations

Mutation	Gene	Mutated sequence used for vaccination	Substitution (WT, AA#, Mut)	Reactive T cell subtype	MHC I score (best prediction)	Response after vaccination with	
						Peptide	RNA
B16-M05	Eef2	FVVKAYLPVNESFAFTADLRSNTGGQA	G795A	CD4+	1.1	x	
B16-M08	Ddx23	ANFESGKHKYRQTAMFTATMPPAVERL	V602A	CD4+	1.3		x
B16-M12	Gnas	TPPPEEAMPFEFNGPAQGDHSQPPLQV	S111G	CD4+	1.2	x	
B16-M17	Tnpo3	VVDRNPQFLDPVLAYLMKGLCEKPLAS	G504A	CD4+	1.0	x	
B16-M20	Tubb3	FRRKAFLHWYTGEAMDEMEFTEAESNM	G402A	CD4+	1.9	x	
B16-M21	Atp11a	SSPDEVALVEGVQSLGFTYLRLKDNYM	R552S	CD4+	0.1		x
B16-M22	Asf1b	PKPDFSQLQRNILPSNPRVTRFHINWD	A141P	CD4+	1.7	x	
B16-M24	Dag1	TAVITPPTTTTKKARVSTPKPATPSTD	P425A	CD4+	2.2	x	
B16-M25	Plod1	STANYNTSHLNNDVWQIFENPVDWKEK	F530V	CD4+	0.1	x	x
B16-M27	Obsl1	REGVELCPGNKYEMRRHGTTHSLVIHD	T1764M	CD8+	2.3	x	x
B16-M28	Ppp1r7	NIEGIDKLTQLKKPFLVNNKINKIENI	L170P	CD4+	3.2	x	x
B16-M29	Mthfd1l	IPSGTTILNCFHDVLSGKLSGGSPGVP	F294V	CD4+	1.7	x	
B16-M30	Kif18b	PSKPSFQEFVDWENVSPELNSTDQPFL	K739N	CD4+	1.2	x	x
B16-M33	Pbk	DSGSPFPAAVILRDALHMARGLKYLHQ	V145D	CD8+	0.1		x
B16-M36	Tm9sf3	CGTAFFINFIAIYHHASRAIPFGTMVA	Y382H	CD4+	0.2	x	
B16-M44	Cpsf3l	EFKHIKAFDRTFANNPGPMVVFATPGM	D314N	CD4+	0.5	x	x
B16-M45	Mkrn1	ECRITSNFVIPSEYWVEEKEEKQKLIQ	N346Y	CD4+	1.4	x	
B16-M46	Actn4	NHSGLVTFQAFIDVMSRETTDTDTADQ	F835V	CD4+	0.2	x	x
B16-M47	Rpl13a	GRGHLLGRLAAIVGKQVLLGRKVVVVR	A24G	CD4+	0.5		x
B16-M48	Def8	SHCHWNDLAVIPAGVVHNWDFEPRKVS	R255G	CD4+	3.8	x	x
B16-M50	Sema3b	GFSQPLRRLVLHVVSAAQAERLARAEE	L663V	CD4+	2.9	x	x

B16F10 mutations determined to be immunogenic upon peptide or RNA immunization (see Fig. 1 and Extended Data Fig. 1). WT, wild type; AA#, position of mutated amino acid; Mut, mutation.

Extended Data Table 2 | Immunogenic CT26 mutations

Mutation	Gene	Mutated sequence used for vaccination	Substitution (WT, AA#, Mut)	Reactive T cell subtype	MHC I score (best prediction)
CT26-M03	Slc20a1	DKPLRRNNSYTSYIMAICGMPLDSFRA	T425I	CD4+	0,3
CT26-M12	Gpc1	YRGANLHLEETLAGFWARLLERLFKQL	E165G	CD8+	1,9
CT26-M13	Nphp3	AGTQCEYWASRALDSEHSIGSMIQLPQ	G234D	CD4+	0,1
CT26-M19	Tmem87a	QAIVRGCSMPGPWRSGRLLVSRRWSVE	G63R	CD8+	0,7
CT26-M20	Slc4a3	PLLPFYPPDEALEIGLELNSSALPPTE	T373I	CD4+	0,9
CT26-M24	Cxcr7	MKAFIFKYSAKTGFTKLIDASRVSETE	L340F	CD4+	1,8
CT26-M26	E2f8	VILPQAPSGPSYATYLQPAQAQMLTPP	I522T	CD8+	0,1
CT26-M27	Agxt2l2	EHIHRAGGLFVADAIQVGFGRIGKHFW	E247A	CD4+	0,2
CT26-M35	Nap1l4	HTPSSYIETLPKAIKRRINALKQLQVR	V63I	CD4+	0,7
CT26-M37	Dhx35	EVIQTSKYYMRDVIAIESAWLLELAPH	T646I	CD4+	0,1
CT26-M39	Als2	GYISRVTAGKDSYIALVDKNIMGYIAS	L675I	CD8+	0,2
CT26-M42	Deptor	SHDSRKSTSFMSVNPSKEIKIVSAVRR	S253N	CD4+	0,3
CT26-M43	Tdg	AAYKGHHYPGPGNYFWKCLFMSGLSEV	H169Y	CD4+	0,3
CT26-M55	Dkk2	EGDPCLRSSDCIDEFCCARHFWTKICK	G192E	CD4+	9,7
CT26-M58	Rpap2	CGYPLCQKKLGVISKQKYRISTKTNKV	P113S	CD4+	11,3
CT26-M68	Steap2	VTSIPSVSNALNWKEFSFIQSTLGYVA	R388K	CD4+	6,8
CT26-M75	Usp26	KTTLSHTQDSSQSLQSSSDSSKSSRCS	S715L	n.d.	5,8
CT26-M78	Nbea	PAPRAVLTGHDHEIVCVSVCAELGLVI	V576I	CD4+	6,3
CT26-M90	Aldh18a1	LHSGQNHLKEMAISVLEARACAAAGQS	P154S	CD4+	8,3
CT26-M91	Zc3h14	NCKYDTKCTKADCLFTHMSRRASILTP	P497L	CD4+	8,8
CT26-M93	Drosha	LRSSLVNNRTQAKIAEELGMQEYAITN	V1189I	CD4+	9,9

CT26 mutations determined to be immunogenic upon RNA immunization (see Fig. 1). WT, wild type; AA#, position of mutated amino acid; Mut, mutation.

Extended Data Table 3 | Immunogenic 4T1 mutations

Mutation	Gene	Mutated sequence used for vaccination	Substitution (WT, AA#, Mut)	Reactive T cell subtype
4T1-M2	Gen1	IPHNPRVAVKTTNNLVMKNSVCLERDS	K707N	CD4
4T1-M3	Polr2a	LAAQSLGEPATQITLNTFHYAGVSAKN	M1102I	CD4
4T1-M8	Tmtc2	QGVTVLAVSAVYDIFVFHRLKMKQILP	V201I	CD8
4T1-M14	Zfr	AHIRGAKHQKVVTLHTKLGKPIPSTEP	K411T	CD4
4T1-M16	Cep120	ELAWEIDRKVLHQNRLQRTPIKLQCFA	H68N	CD4
4T1-M17	Malt1	FLKDRLLEDKKIAVLLDEVAEDMGKCH	T534A	CD4
4T1-M20	Wdr11	NDEPDLDPVQELIYDLRSQCDAIRVTK	T340I	CD8
4T1-M22	Kbtbd2	DAAALQMIIAYAYRGNLAVNDSTVEQL	T91R	CD4
4T1-M25	Adamts9	KDYTAAGFSSFQKLRLDLTSMQIITTD	I623L	CD4
4T1-M26	Pzp	AVKEEDSLHWQRPEDVQKVKALSFYQP	G1199E	CD8
4T1-M27	Gprc5a	FAICFSCLLAHALNLIKLVRGRKPLSW	F119L	CD8
4T1-M30	Enho	MGAAISQGAIIAIVCNGLVGFLL	L10I	CD4
4T1-M31	Dmrta2	EKYPRTPKCARCGNHGVVSALKGHKRY	R73G	CD4
4T1-M32	Rragd	SHRSCSHQTSAPSPKALAHNGTPRNAI	L268P	CD4
4T1-M35	Zzz3	KELLQFKKLKKQNLQQMQAESGFVQHV	K311N	CD8
4T1-M39	Ilkap	RKGEREEMQDAHVSLNDITQECNPPSS	I27S	CD4
4T1-M40	Cenpf	RVEKLQLESELNESRTECITATSQMTA	D1327E	CD4

4T1 mutations determined to be immunogenic upon RNA immunization (see Fig. 1). WT, wild type; AA#, position of mutated amino acid; Mut, mutation.

Extended Data Table 4 | CT26 mutated epitopes encoded in pentatope 1+2

Pentatope	Mutation	Gene	Mutated sequence used for vaccination	Substitution (WT, AA#, Mut)	Reactive T cell subtype
1	CT26-M19	Tmem87a	QAIVRGCSMPGPWRSGRLLVSRRWSVE	G63R	CD8+
1	CT26-M39	Als2	GYISRVTAGKDSYIALVDKNIMGYIAS	L675I	CD8+
1	CT26-M13	Nphp3	AGTQCEYWASRALDSEHSIGSMIQLPQ	G234D	CD4+
1	CT26-M55	Dkk2	EGDPCLRSSDCIDEFCCARHFWTKICK	G192E	CD4+
1	CT26-M68	Steap2	VTSIPSVSNALNWKEFSFIQSTLGYVA	R388K	CD4+
2	CT26-M20	Slc4a3	PLLPFYPPDEALEIGLELNSSALPPTE	T373I	CD4+
2	CT26-M26	E2f8	VILPQAPSGPSYATYLQPAQAQMLTPP	I522T	CD8+
2	CT26-M03	Slc20a1	DKPLRRNNSYTSYIMAICGMPLDSFRA	T425I	CD4+
2	CT26-M37	Dhx35	EVIQTSKYYMRDVIAIESAWLLELAPH	T646I	CD4+
2	CT26-M27	Agxt2l2	EHIHRAGGLFVADAIQVGFGRIGKHFW	E247A	CD4+

Ten immunogenic mutated epitopes were used for generation of two pentatopes used for therapeutic vaccination shown in Fig. 3b. WT, wild type; AA#, position of mutated amino acid; Mut, mutation.

Extended Data Table 5 | *In silico* prediction of CT26 mutations with abundant expression and favourable MHC class II binding properties

Mutation	Gene	Mutated sequence used for vaccination	Substitution (WT, AA#, Mut)	Expression (NVRC)	MHC II score (best prediciton)
CT26-E1	Asns	DSVVIFSGEGSDEFTQGYIYFHKAPSP	L370F	1428,05	45,45
CT26-E2	Cd34	PQTSPTGILPTTSNSISTSEMTWKSSL	D120N	1150,85	23,76
CT26-E3	Actb	WIGGSILASLSTFHQMWISKQEYDESG	Q353H	974,16	8,30
CT26-E4	Tmbim6	SALGSLALMIWLMTTPHSHETEQKRLG	A73T	825,51	2,96
CT26-E5	Glud1	DLRTAAYVNAIEKIFKVYNEAGVTFT	V546I	619,54	8,04
CT26-E16	Eif4g2	KLCLELLNVGVESNLILKGVILLIVDK	K108N	327,79	20,99
CT26-E17	Sept7	NVHYENYRSRKLATVTYNGVDNNKNKG	A314T	316,98	6,47
CT26-E18	Fn1	YTVSVVALHDDMENQPLIGIQSTAIPA	S1710N	303,62	17,41
CT26-E19	Brd2	KPSTLRELERYVLACLRKKPRKPYTIR	S703A	301,83	7,86
CT26-E20	Uchl3	KFMERDPDELRFNTIALSAA	A224T	301,78	9,75
CT26-ME1	Aldh18a1	LHSGQNHLKEMAISVLEARACAAAGQS	P154S	67,73	0,05
CT26-ME2	Ubqln1	DTLSAMSNPRAMQVLLQIQQGLQTLAT	A62V	84,08	0,24
CT26-ME3	Ppp6r1	DGQLELLAQGALDNALSSMGALHALRP	D309N	139,80	0,44
CT26-ME4	Trip12	WKGGPVKIDPLALMQAIERYLVVRGYG	V1328M	83,09	0,49
CT26-ME5	Pcdhgc3	QDINDNNPSFPTGKMKLEISEALAPGT	E139K	86,16	0,54
CT26-ME6	Cad	SDPRAAYFRQAENDMYIRMALLATVLG	G2139D	152,86	0,55
CT26-ME7	Smarcd1	MDLLAFERKLDQTVMRKRLDIQEALKR	I161V	125,85	0,60
CT26-ME8	Ddx27	ITTCLAVGGLDVKFQEAALRAAPDILI	S297F	61,82	0,62
CT26-ME9	Snx5	KARLKSKDVKLAEAHQQECCQKFEQLS	T341A	120,27	0,73
CT26-ME10	Lin7c	GEVPPQKLQALQRALQSEFCNAVREVY	V41A	71,24	1,09

CT26 mutations selected for high expression with (ME) or without (E) consideration of the MHC II percentile rank (IEDB consensus version 2.5). WT, wild type; AA#, position of mutated amino acid; Mut, mutation.

LETTER

TP53 loss creates therapeutic vulnerability in colorectal cancer

Yunhua Liu[1], Xinna Zhang[2,3], Cecil Han[1], Guohui Wan[1], Xingxu Huang[4], Cristina Ivan[2,3], Dahai Jiang[2,3], Cristian Rodriguez-Aguayo[3,5], Gabriel Lopez-Berestein[3,5], Pulivarthi H. Rao[6], Dipen M. Maru[7], Andreas Pahl[8], Xiaoming He[9], Anil K. Sood[1,2,3], Lee M. Ellis[10], Jan Anderl[8] & Xiongbin Lu[1,3]

TP53*, a well-known tumour suppressor gene that encodes p53, is frequently inactivated by mutation or deletion in most human tumours[1,2]. A tremendous effort has been made to restore p53 activity in cancer therapies[3–7]. However, no effective p53-based therapy has been successfully translated into clinical cancer treatment owing to the complexity of p53 signalling. Here we demonstrate that genomic deletion of *TP53* frequently encompasses essential neighbouring genes, rendering cancer cells with hemizygous *TP53* deletion vulnerable to further suppression of such genes. *POLR2A* is identified as such a gene that is almost always co-deleted with *TP53* in human cancers. It encodes the largest and catalytic subunit of the RNA polymerase II complex, which is specifically inhibited by α-amanitin[8,9]. Our analysis of The Cancer Genome Atlas (TCGA) and Cancer Cell Line Encyclopedia (CCLE) databases reveals that POLR2A expression levels are tightly correlated with its gene copy numbers in human colorectal cancer. Suppression of POLR2A with α-amanitin or small interfering RNAs selectively inhibits the proliferation, survival and tumorigenic potential of colorectal cancer cells with hemizygous *TP53* loss in a p53-independent manner. Previous clinical applications of α-amanitin have been limited owing to its liver toxicity[10]. However, we found that α-amanitin-based antibody–drug conjugates are highly effective therapeutic agents with reduced toxicity[11]. Here we show that low doses of α-amanitin-conjugated anti-epithelial cell adhesion molecule (EpCAM) antibody lead to complete tumour regression in mouse models of human colorectal cancer with hemizygous deletion of *POLR2A*. We anticipate that inhibiting POLR2A will be a new therapeutic approach for human cancers containing such common genomic alterations.

Genomic deletion of a tumour suppressor gene often encompasses several neighbouring genes that may not contribute to cancer development, but are essential for cell proliferation and survival[12]. This partial loss of essential genes has been postulated to render cancer cells highly vulnerable to further inhibition of those genes[13,14]. Analysis of TCGA revealed that hemizygous deletion of the *TP53* gene occurs frequently in human cancers (Fig. 1a). We identified *POLR2A* as an essential gene in the proximity of *TP53* (Fig. 1b). Concomitant deletion of *POLR2A* occurs in virtually all the human colorectal tumours containing hemizygous *TP53* deletion (Fig. 1c).

Among the 12 subunits in the human RNA polymerase II complex, *POLR2A* encodes the largest subunit that is indispensable for the polymerase activity in messenger RNA synthesis. Inhibiting POLR2A with a specific inhibitor, α-amanitin, causes extensive cell death, and homozygous deletion of *POLR2A* is lethal in human cells[9,15]. We found that 104 (53%) out of 195 colorectal cancer (CRC) cases bear hemizygous loss of the 17p13 region, resulting in concomitant deletion of *TP53* and

POLR2A (Fig. 1c). However, no homozygous deletion of *POLR2A* was observed, consistent with the notion that *POLR2A* is essential for cell survival. Analysis of TCGA and CCLE databases revealed that expression of *POLR2A* is tightly correlated with its gene copy number (Fig. 1d). This positive correlation was also validated in 20 pairs of matched normal and CRC tissue samples and human CRC tissue microarray (Fig. 1e and Extended Data Fig. 1). POLR2A[loss] (hemizygous loss) cell lines expressed POLR2A proteins at significantly lower levels than POLR2A[neutral] cell lines (Fig. 1f–h). Unlike POLR2A, p53 levels are determined by post-transcriptional and post-translational events[16]. Despite a correlation between *TP53* copy number and mRNA expression, p53 protein levels are not associated with its gene copy numbers in human CRC (Extended Data Fig. 2 and Fig. 1h).

To assess the sensitivity of cells to POLR2A inhibition, a panel of POLR2A[neutral] (HCT116, SW480) and POLR2A[loss] (SW837, SNU283) cells were treated with α-amanitin. Treatment of α-amanitin at high concentrations ($\geq 1\,\mu g\,ml^{-1}$) caused complete cell death in all four cell lines. However, at concentrations ranging from 0 to $1.0\,\mu g\,ml^{-1}$, α-amanitin inhibition had significantly higher levels of cell-killing effect on the POLR2A[loss] cells than on the POLR2A[neutral] cells (Fig. 2a, b). The half-maximum inhibitory concentration (IC$_{50}$) was ~$1.0\,\mu g\,ml^{-1}$ for the POLR2A[neutral] cells, which was tenfold greater than that of the POLR2A[loss] cells. By contrast, the POLR2A[loss] cells did not show any greater sensitivity to the treatment of actinomycin D, a nonspecific transcription inhibitor (Extended Data Fig. 3a). In direct competition assays, the POLR2A[neutral] cells (HCT116, SW480) stably expressing *POLR2A* short hairpin RNAs (shRNAs) only had modestly reduced proliferation, in comparison with that of the corresponding cells expressing control shRNAs (Extended Data Fig. 3b, c). However, silencing POLR2A in the POLR2A[loss] cells (SNU283, SW837) led to markedly reduced proliferation. We generated HCT116 and SNU283 cell lines stably expressing doxycycline (Dox)-inducible *POLR2A* shRNAs (Extended Data Fig. 3d). Despite significant knockdown of POLR2A, HCT116 cells continued to proliferate, whereas SNU283 cells exhibited severe G1 cell cycle arrest and apoptosis (Fig. 2c, d and Extended Data Fig. 3e–g). Approximately 50% of the decrease in POLR2A expression (30–100 ng ml^{-1} of Dox) remarkably reduced the proliferation of SNU283 cells, but only had a modest effect on HCT116 cells (Fig. 2d). Results of rescue experiments demonstrated that gradual re-expression of exogenous POLR2A in SNU283 and SW837 cells restored their resistance to α-amanitin up to a level comparable to that of the POLR2A[neutral] cells (Fig. 2e, f and Extended Data Fig. 4).

To exclude genetic difference across cell lines, we used the CRISPR (clustered regularly interspaced short palindromic repeat)/Cas9 system to generate an isogenic HCT116 cell line with hemizygous loss of

Figure 1 | Expression of *POLR2A*, but not *TP53*, is correlated with the gene copy number. a, Frequencies of hemizygous deletion of *TP53* in human cancers. AML, acute myeloid leukaemia; DLBC, diffuse large B-cell lymphoma; GBM, glioblastoma. **b**, Schematic diagram of genes adjacent to *TP53* in human genome. **c**, Concomitant deletion of *POLR2A* in human CRCs containing hemizygous loss of *TP53*. **d**, Scatterplots of *POLR2A* copy number versus mRNA expression in TCGA and CCLE databases. Pearson correlation coefficients (*r*) and *P* values are shown. **e**, POLR2A protein levels in matched normal and CRC tissue samples. Error bars denote s.d. **f**, **g**, Copy numbers (**f**) and relative mRNA expression levels (**g**) of *POLR2A* in human CRC cell lines. Data are mean and s.d. of three independent experiments. **h**, Protein levels of POLR2A and p53 in human CRC cell lines.

POLR2A (refs 17, 18) (Extended Data Fig. 5a–c). POLR2A expression was significantly reduced in the POLR2A^{loss} HCT116 cells (Extended Data Fig. 5d). The POLR2A^{loss} and parental HCT116 cells exhibited similar proliferation rates (Extended Data Fig. 5e), indicating that one allele of *POLR2A* is sufficient to maintain cell proliferation. However, hemizygous deletion of POLR2A markedly sensitized HCT116 cells to α-amanitin treatment with an IC$_{50}$ value of ~0.1 μg ml^{-1}, which is eightfold lower than that of parental HCT116 cells (Fig. 2g, h). As a control, no substantial difference on their sensitivity to actinomycin D was observed (Extended Data Fig. 5f). Knockdown of POLR2A resulted in profoundly reduced proliferation in the POLR2A^{loss} cells, but not in the isogenic POLR2A^{neutral} cells (Extended Data Fig. 5g). In the isogenic pair of HCT116 cell lines expressing Dox-inducible *POLR2A* shRNAs, a low-dose of Dox (100 ng ml^{-1}) was sufficient to remarkably kill the POLR2A^{loss} cells, while high doses of Dox only had minimal effects on the POLR2A^{neutral} cells (Fig. 2i, j and Extended Data Fig. 5h).

Blockage of RNA polymerase may lead to p53 accumulation and activation[19]. To examine the effects of p53, we recapitulated the concomitant deletion of *TP53* and *POLR2A* in HCT116 and xenografted human primary CRC (xhCRC) cell lines (Fig. 3a and Extended Data Fig. 6a–c). The xhCRC cell line (*TP53*$^{+/+}$ *POLR2A*$^{+/+}$), established from a freshly isolated xenografted human colon tumour, demonstrated enhanced tumorigenicity *in vivo*[20]. Except for slightly increased cell proliferation, no significant changes in their sensitivity to α-amanitin were observed in either xhCRC or HCT116 cells with hemizygous deletion of *TP53*. By contrast, hemizygous loss of *POLR2A* markedly sensitized these cells to α-amanitin treatment regardless of their *TP53* status (Fig. 3b and Extended Data Fig. 6d–g). The mRNA synthesis activity of the RNA polymerase II is essential to any type of cells including therapy-resistant tumour cells. We examined the drug sensitivity of POLR2A^{neutral} and POLR2A^{loss} cells to three major chemotherapy drugs for CRC: 5-fluorouracil (5-FU), oxaliplatin and 7-ethyl-10-hydroxy-camptothecin (SN-38). Inhibition of POLR2A by α-amanitin significantly enhanced the cell-killing effects of all three drugs in the POLR2A^{loss} xhCRC cells, but had no notable effects on the POLR2A^{neutral} cells (Fig. 3c), suggesting therapeutic vulnerability of POLR2A^{loss} colorectal tumours. Free α-amanitin is toxic to liver owing to its interaction with OATP1B3, a transporter exclusively expressed on the membrane of hepatocytes[10]. However, when conjugated with specific antibodies, α-amanitin is no longer a substrate for OATP1B3 (refs 10, 11, 21). This strategy overcomes the toxicity of α-amanitin for clinical applications. We used α-amanitin conjugated to a monoclonal antibody (HEA125) against EpCAM, a cancer antigen overexpressed in adenocarcinomas[11,22]. The ama–HEA125 conjugate selectively killed the POLR2A^{loss} xhCRC and HCT116 cancer cells in a p53-independent manner, and reduced the effective doses of α-amanitin by at least 10,000-fold (IC$_{50}$ ~ 0.01 ng ml^{-1}) *in vitro* (Fig. 3d and Extended Data Fig. 6h).

To test the anti-tumour effect of POLR2A inhibition *in vivo*, HCT116 and SNU283 cells expressing Dox-inducible *POLR2A* shRNA were injected subcutaneously into non-obese diabetic/severe
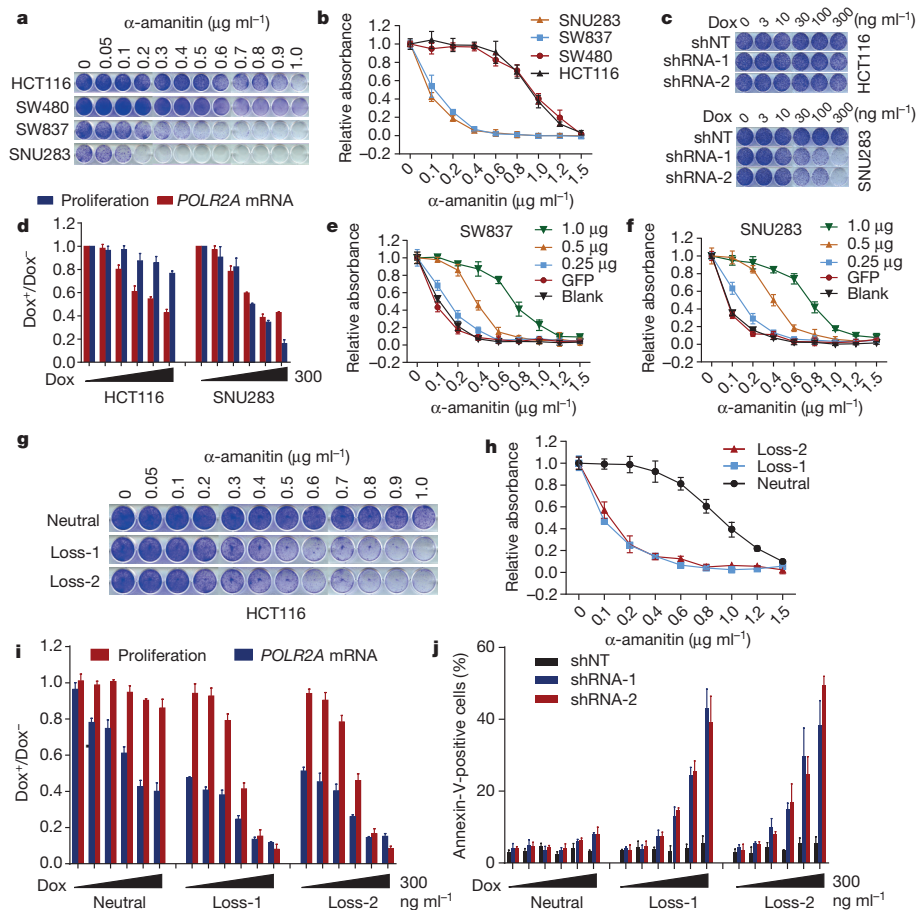
Figure 2 | POLR2A^loss cells are highly sensitive to the POLR2A inhibition. a, b, POLR2A^loss cells (SW837, SNU283) are significantly more sensitive to α-amanitin treatment than POLR2A^neutral cells (HCT116, SW480). Crystal violet staining of cells (**a**) and quantification analyses (**b**) are shown. **c,** Dox-induced suppression of POLR2A inhibited the proliferation of SNU283 cells, but not of HCT116 cells. shNT denotes non-targeting control shRNA. **d,** Correlation between *POLR2A* mRNA expression and cell proliferation in HCT116 and SNU283 cells expressing Dox-inducible *POLR2A* shRNA. **e, f,** Survival curves of SW837 (**e**) and SNU283 (**f**) cells treated with α-amanitin after transfection with increasing amounts of *PORL2A* expression vector DNA. **g, h,** POLR2A^loss HCT116 cells are significantly more sensitive to α-amanitin treatment than the parental POLR2A^neutral cells. Crystal violet staining of cells (**g**) and quantification analyses (**h**) are shown. **i, j,** Correlation between *POLR2A* mRNA expression and cell proliferation (**i**) or apoptosis (**j**) in POLR2A^neutral and POLR2A^loss HCT116 cells. Data are mean and s.d. of three independent experiments.
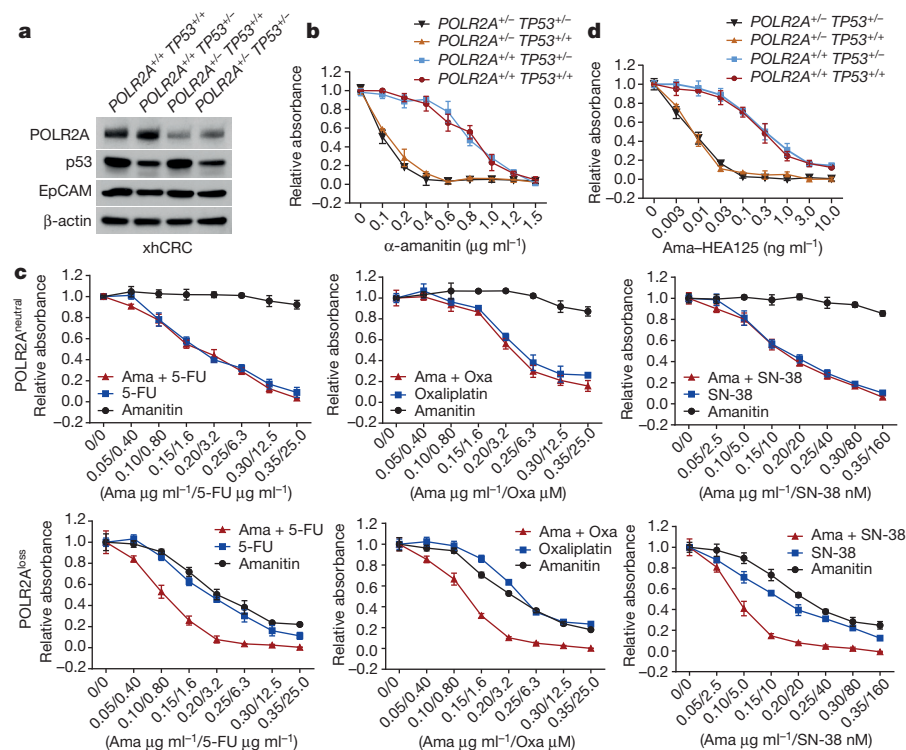
Figure 3 | The sensitivity of POLR2A^loss cells to POLR2A inhibition is independent of p53. a, Protein levels of POLR2A, p53 and EpCAM in a panel of isogenic human xhCRC cell lines. **b,** Cell proliferation of isogenic xhCRC cells treated with α-amanitin. **c,** Sensitivity of isogenic xhCRC cells to 5-FU, oxaliplatin (Oxa) or SN-38 treatment combined with or without α-amanitin (Ama). **d,** Cell proliferation of isogenic xhCRC cells treated with ama–HEA125. Data are mean and s.d. of three independent experiments.
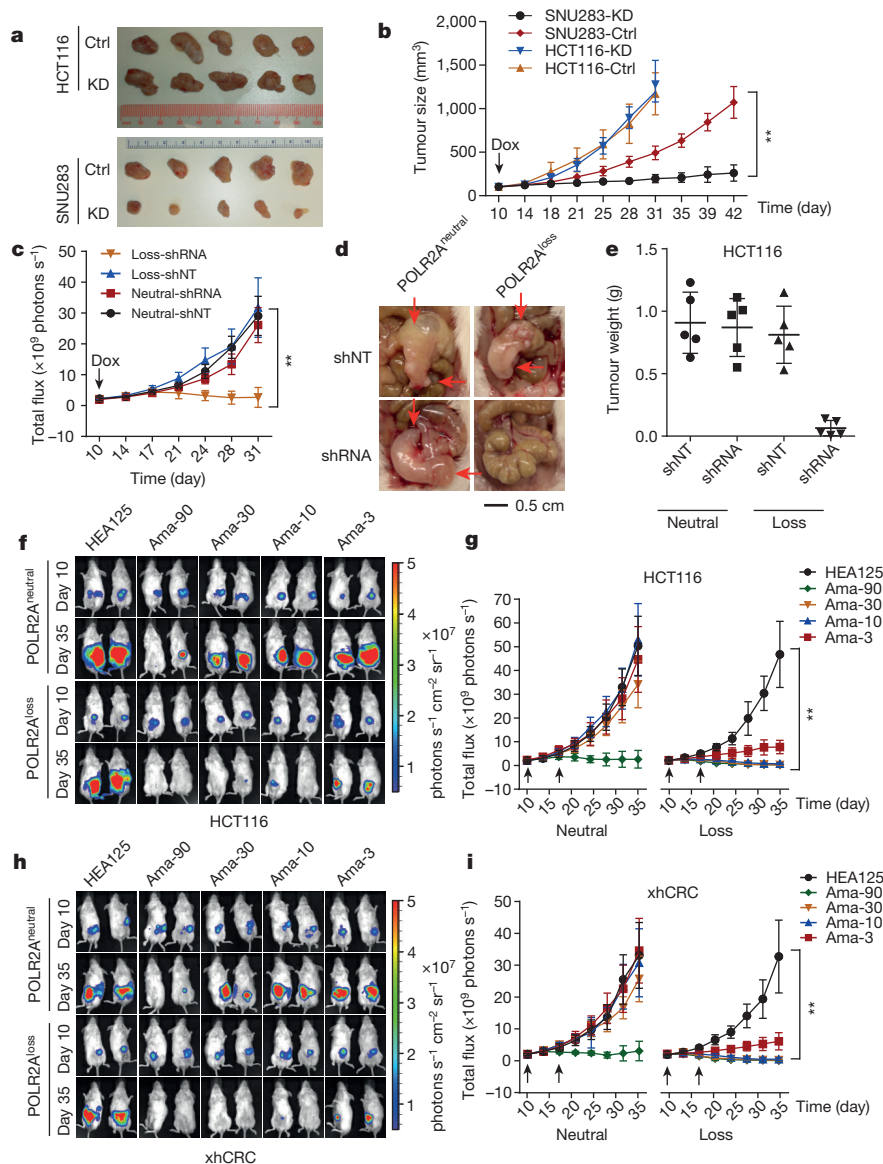
Figure 4 | Suppression of POLR2A selectively inhibits the POLR2A^{loss} tumour growth. a, b, Gross tumour images (**a**) and growth curves (**b**) of xenograft tumours derived from subcutaneously implanted HCT116 or SNU283 cells expressing control (ctrl) or Dox-inducible *POLR2A* shRNA. KD, knockdown. $n = 5$ mice per group. Error bars denote s.e.m. **c–e,** Tumour growth curves (**c**, **P < 0.01, error bars denote s.e.m.), gross tumour images (**d**) and weights (**e**, error bars denote s.d.) of xenograft tumours derived from orthotopically implanted POLR2A^{neutral} and POLR2A^{loss} HCT116 cells expressing Dox-inducible control or *POLR2A* shRNA (shNT or shRNA, respectively). $n = 5$ mice per group. **f–i,** Representative bioluminescent images (**f, h**) and tumour growth curves (**g, i**) of orthotopic xenograft tumours derived from POLR2A^{neutral} and POLR2A^{loss} HCT116 (**f, g**) or xhCRC cells (**h, i**) that received dual intraperitoneal injections of HEA125 antibody or ama–HEA125 antibody–drug conjugate (3, 10, 30 and 90 µg kg^{-1}). $n = 10$ mice per group. Error bars denote s.e.m.

combined immunodeficiency (NOD/SCID) mice. After initial tumour establishment, administration of Dox (1.0 µg ml^{-1}) in drinking water suppressed POLR2A expression and consequently inhibited the growth of SNU283-derived tumours (Fig. 4a, b and Extended Data Fig. 7a). However, no substantial differences were observed between control and POLR2A-knockdown HCT116-derived tumours. Histopathological analyses demonstrated that POLR2A-knockdown SNU283 tumours had significantly reduced cell proliferation, but more apoptotic cells, as compared with the corresponding control tumours (Extended Data Fig. 7b, c). By contrast, no significant changes were observed in the control or POLR2A-knockdown HCT116 tumours. However, hemizygous deletion of POLR2A sensitized HCT116-derived tumours to Dox treatment (Extended Data Fig. 7d, e). Next, we used an orthotopic tumour model by injecting POLR2A^{neutral} and POLR2A^{loss}

HCT116 cells into the caecal wall of NOD/SCID mice. *In vivo* tumour imaging demonstrated that Dox-induced POLR2A inhibition led to a significant decrease in tumour growth kinetics in the POLR2A^{loss} tumours, but not in the control POLR2A^{neutral} tumours (Fig. 4c–e and Extended Data Fig. 7f). To test the efficacy of POLR2A silencing *in vivo* further, we used a nanoliposomal delivery platform, DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine), for systemic delivery of *POLR2A* siRNAs[23] (Extended Data Fig. 8a, b). After four weeks of systemic therapy, compared with control short interfering RNA (siRNA) and DOPC treatment, mice in the 125 µg kg^{-1} of *POLR2A*-siRNA–DOPC treatment groups had pronounced growth reduction of POLR2A^{loss} tumours, while POLR2A^{neutral} tumours only had significant growth inhibition even at the dose of 1,000 µg kg^{-1} (Extended Data Fig. 8c–h).

Next, we investigated the anti-tumour activity of ama–HEA125 in orthotopic tumour models established by isogenic pairs of xhCRC and HCT116 cells (TP53/POLR2A^{neutral} and TP53/POLR2A^{loss}) (Fig. 4f–i and Extended Data Fig. 9a–c). In a dose-escalation experiment, ama–HEA125 was administered to the tumour-bearing mice as dual intraperitoneal injections at the dose of 3, 10, 30 or 90 μg with respect to α-amanitin per kilograms of mouse body weight. In the mice bearing POLR2A^{loss} tumours, control HEA125-treated mice showed continuous tumour growth within 25 days after antibody injection. The treatment of ama–HEA125 showed strong inhibition on tumour growth even at the lowest dose of 3 μg kg^{-1}. All the POLR2A^{loss} tumours responded to ama–HEA125 treatment, and tumour volume regressed markedly (Fig. 4f–i). Complete tumour regression was observed in 10 out of 10 (90 μg kg^{-1}), 8 out of 10 (30 μg kg^{-1}) and 6 out of 10 (10 μg kg^{-1}) mice bearing POLR2A^{loss} HCT116 tumours 25 days after the first administration of ama–HEA125 (Fig. 4f, g). By contrast, significant tumour inhibition was observed in the mice bearing POLR2A^{neutral} tumours only at the highest dose of 90 μg kg^{-1}, but not at the doses of 3–30 μg kg^{-1}. Similar results were observed in the mice bearing xhCRC-derived tumours (Fig. 4h, i). Consistent with our previous study[11], treatment of ama–HEA125 at the tested doses had no notable toxicity *in vivo*. Analysis of body weights and blood liver enzymes did not reveal any substantial differences between the ama–HEA125-treated group and the control HEA125-treated group (Extended Data Fig. 9d, e), suggesting negligible systemic toxicity of the ama–HEA125 conjugate. In addition, we also tested the anti-tumour activity of ama–HEA125 in orthotopic tumours derived from the POLR2A^{neutral} SW480 cells and the POLR2A^{loss} SW837 cells (Extended Data Fig. 10). Treatment of ama–HEA125 at the 10 μg kg^{-1} dose was sufficient to inhibit tumour growth in all the POLR2A^{loss} tumours, while POLR2A^{neutral} (SW480) or POLR2A-restored (SW837 expressing exogenous POLR2A) tumours were only significantly inhibited by ama–HEA125 at the dose of 90 μg kg^{-1}.

Much effort has been directed towards restoring p53 activity in human cancer cells[3,4,7,24]. Instead of pharmacological intervention of p53 signalling, the principle of collateral vulnerability to POLR2A inhibition provides a novel strategy for cancer therapy. Our preclinical studies with α-amanitin demonstrate therapeutic efficacy of the strategy for treating CRC. Given that hemizygous loss of *TP53* is a frequent genomic event in many human cancers, the therapeutic strategy described here may have broad applications. While clinical applications of α-amanitin are limited by its liver toxicity, conjugation of α-amanitin to antibodies against tumour-specific biomarkers minimizes its systemic toxicity and, moreover, increases the efficacy of treatment. The identification of new pharmacological agents that target POLR2A should significantly impact personalized cancer therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Petitjean, A. *et al.* Impact of mutant p53 functional properties on *TP53* mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* **28**, 622–629 (2007).
2. Vazquez, A. *et al.* The genetics of the p53 pathway, apoptosis and cancer therapy. *Nature Rev. Drug Discov.* **7**, 979–987 (2008).
3. Cheok, C. F. *et al.* Translating p53 into the clinic. *Nature Rev. Clin. Oncol.* **8**, 25–37 (2011).
4. Lane, D. P., Cheok, C. F. & Lain, S. p53-based cancer therapy. *Cold Spring Harb. Perspect. Biol.* **2**, a001222 (2010).
5. Chène, P. Inhibiting the p53–MDM2 interaction: an important target for cancer therapy. *Nature Rev. Cancer* **3**, 102–109 (2003).
6. Wade, M., Li, Y. C. & Wahl, G. M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nature Rev. Cancer* **13**, 83–96 (2013).
7. Haupt, S. & Haupt, Y. Manipulation of the tumor suppressor p53 for potentiating cancer therapy. *Semin. Cancer Biol.* **14**, 244–252 (2004).
8. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108 (2011).
9. Lindell, T. J. *et al.* Specific inhibition of nuclear RNA polymerase II by α-amanitin. *Science* **170**, 447–449 (1970).
10. Letschert, K. *et al.* Molecular characterization and inhibition of amanitin uptake into human hepatocytes. *Toxicol. Sci.* **91**, 140–149 (2006).
11. Moldenhauer, G. *et al.* Therapeutic potential of amanitin-conjugated anti-epithelial cell adhesion molecule monoclonal antibody against pancreatic carcinoma. *J. Natl. Cancer Inst.* **104**, 622–634 (2012).
12. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability–an evolving hallmark of cancer. *Nature Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
13. Nijhawan, D. *et al.* Cancer vulnerabilities unveiled by genomic loss. *Cell* **150**, 842–854 (2012).
14. Muller, F. L. *et al.* Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* **488**, 337–342 (2012).
15. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
16. Toledo, F. & Wahl, G. M. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nature Rev. Cancer* **6**, 909–923 (2006).
17. Wang, T. *et al.* Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
18. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
19. Derheimer, F. A. *et al.* RPA and ATR link transcriptional stress to p53. *Proc. Natl Acad. Sci. USA* **104**, 12778–12783 (2007).
20. Lu, J. *et al.* Endothelial cells promote the colorectal cancer stem cell phenotype through a soluble form of Jagged-1. *Cancer Cell* **23**, 171–185 (2013).
21. Faulstich, H. & Fiume, L. Protein conjugates of fungal toxins. *Methods Enzymol.* **112**, 225–237 (1985).
22. Went, P. T. *et al.* Frequent EpCam protein expression in human carcinomas. *Hum. Pathol.* **35**, 122–128 (2004).
23. Pecot, C. V. *et al.* Therapeutic silencing of KRAS using systemically delivered siRNAs. *Mol. Cancer Ther.* **13**, 2876–2885 (2014).
24. Goldstein, I. *et al.* Understanding wild-type and mutant p53 activities in human cancer: new landmarks on the way to targeted therapies. *Cancer Gene Ther.* **18**, 2–11 (2011).

Author Contributions Y.L., X.Z. and X.L. directed the work and wrote the manuscript. C.H. and G.W. generated vectors and performed cell-based assays. X.H. provided technical guidance on CRISPR/Cas9 and gene knockout experiments. D.M.M. provided clinical tissue samples and technical guidance. D.J., C.R.-A. and G.L.-B. produced siRNA–DOPC. P.H.R. carried out FISH experiments. J.A. and A.P. provided α-amanitin-conjugated antibodies and technical guidance in xenograft tumour experiments. L.M.E., X.H., C.I. and A.K.S. helped to analyse data. All authors discussed the results and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X. L. (xlu2@mdanderson.org).

METHODS

Cell culture, antibodies and western blot analysis. HCT116, SW480, SW837, HT29, DLD1 and HT29 cell lines were obtained from the American Type Culture Collection and cultured under standard conditions specified by the manufacturer. SNU283 and SNU1197 cell lines were obtained from the Korean Cell Line Bank and cultured in RMPI 1640 medium supplemented with 10% FBS and 2 mmol l^{-1} L-glutamine. Isolation, culture and maintenance of xhCRC cells were performed as previously described[20]. In brief, patient-derived xenografts were collected under sterile conditions and mechanically dissociated, followed by 30 min of incubation in collagenase II at 37 °C. The specimen was filtered through a sterile 100-μm strainer. Red blood cells were eliminated with a hypo-osmotic red blood cell lysis buffer (eBioscience). Mouse cells in xenografted human CRC specimens were removed by negative selection using a mouse MHC class I molecule H-2K anti-body followed by use of a magnetic bead purification kit (Miltenyi). Fibroblasts were removed by negative selection using a MACS magnetic bead separation kit (Miltenyi). The freshly isolated xhCRC cells were maintained on collagen-1-coated culture plates (BD Biosciences) and cultured in MEM supplemented with 10% FBS, vitamins (1×), nonessential amino acids (1×), penicillin-streptomycin (1×), sodium pyruvate (1×) and L-glutamine (1×). All medium supplements were purchased from Sigma.

Anti-POLR2A antibodies were purchased from Santa Cruz (sc-47701) and Abcam (ab140509). Anti-Ki67 antibody (D3B5) and anti-cleaved caspase-3 (Asp175, 5A1E) antibody were obtained from Cell Signalling. Anti-p53 (sc-126), anti-β-actin (sc-1616), horseradish peroxidase (HRP)-anti-goat IgG (sc-2020), HRP-anti-rabbit IgG (sc-2054), and HRP-anti-mouse IgG (sc-2055) antibodies were purchased from Santa Cruz. Cell lysate preparation, SDS–PAGE and western blotting were performed as previously described[25].

RNA isolation and quantitative PCR. Total RNA was isolated using TRIzol reagent (Life Technologies) and then reverse-transcribed using iScript cDNA Synthesis Kit (Bio-Rad). The resulting cDNA was used for qPCR using iTaq Universal SYBR Green Supermix (Bio-Rad) with gene-specific primers and the results were normalized to β-actin control. In reverse transcription PCR (RT–PCR) assays, the primer sequences for TP53 are: 5′-GAGGTTGGCTCTGACTGT ACC-3′ and 5′-TCCGTCCCAGTAGATTACCAC-3′, and for POLR2A are: 5′-TTGTATCCGTACCCACAGCA-3′ and 5′-CATGATCAGCTCCCCATTCT-3′.

shRNA-mediated knockdown of POLR2A. POLR2A-specific shRNA clones were obtained from the MD Anderson shRNA and ORFeome Core Facility (originally from Open Biosystems). Twelve shRNAs targeting POLR2A were screened, of which two shRNAs knocked down protein levels by at least 50% in all four colorectal cancer cell lines tested. Their clone identification numbers and shRNA sequences are V3LHS_645674 (5′-TTAGCTTTGTTCTTCCCGA-3′) and V3LHS_640293 (5′-TGTTGTCCATCTCCTCCCC-3′). The hairpin sequences in the GIPZ vector were cloned into the TRIPZ vector (Dharmacon) using a protocol provided by the manufacturer. The TRIPZ vector is a Dox-inducible system with a red fluorescent protein reporter.

Generation of cells stably expressing Dox-inducible shRNAs. Recombinant lentiviral particles were produced from 293T cells. In brief, 72 μg of shRNA-encoding vector DNA, 54 μg of Delta 8.9 vector DNA and 18 μg of VSVG-encoding vector DNA were transfected into 293T cells (plated in the 245-mm^2 dish) using X-tremeGENE (Roche). Supernatant containing virus particles was collected and filtered 72 h after transfection, concentrated by ultracentrifugation at 90,000g, and resuspended in cell growth medium. To generate stable Dox-inducible cells, HCT116 and SNU283 cells were infected with shRNA-expressing viral particles at a multiplicity of infection (MOI) of 1. Viral solutions were added to cell culture medium containing 4 μg ml^{-1} polybrene. Cells were selected 48 h after infection by 2 μg ml^{-1} puromycin. Single colonies were cultured and propagated, and colonies bearing a single copy of lentiviral DNA insert were identified and analysed for knockdown efficiency.

Direct competition assay using POLR2A shRNA. A single lentiviral copy expressing either shRNA-1 or shRNA-2 was sufficient to suppress POLR2A expression levels in all the four colorectal cell lines tested (Extended Data Fig. 3b). Cancer cells were infected with control or POLR2A shRNA-expressing lentiviruses (pGIPZ backbone expressing green fluorescent protein (GFP)) at the MOI of 2. Two days after infection, GFP-positive cells were sorted using a BD FACSJazz cell sorter (BD Biosciences) at the MD Anderson Flow Cytometry and Cellular Imaging Core. Next, GFP-positive cells were mixed with non-infected and GFP-negative cells at the ratio of 1:1 and cultured for six passages. The numbers of GFP-positive and total cells at each passage were analysed and quantified by flow cytometry and the percentages of GFP-positive cells were calculated.

Generation of sgRNA-expressing vectors and Surveyor assay. Bicistronic expression vector expressing Cas9 and single-guide RNA (sgRNA) was digested with BbsI and treated with alkaline phosphatase, and then the linearized vector was gel-purified for cloning sgRNA-encoding DNA[26]. The pair of oligonucleotide DNA for each sgRNA targeting TP53 or POLR2A was annealed, phosphorylated and ligated to the linearized vector. The sequences of oligonucleotide DNA are listed in Supplementary Table 1. Surveyor assay was performed to test the genome editing efficacy as previously described[27,28]. In brief, cells were seeded into six-well plates at a density of 2×10^5 cells per well. One day after initial seeding, cells were transfected with 2 μg of Cas9/sgRNA-expressing vector DNA and collected 48 h after transfection. Genomic DNA was isolated and a 1-kb DNA fragment containing the sgRNA-targeting site was amplified by high-fidelity PCR and digested by T7 endonuclease I. Genomic DNA isolated from HCT116 cells transfected with control vector DNA was used as control. To allow complementary but mismatched strands to anneal, PCR products were incubated at 95 °C for 5 min, 95 °C to 85 °C at the rate of -2 °C s^{-1} and 85 °C to 25 °C at the rate of -0.1 °C s^{-1}. T7 endonuclease I was added and samples were incubated at 37 °C for 60 min to digest the annealed PCR products at the mismatch sites. T7-endonuclease-I-digested PCR products were analysed by agarose gel electrophoresis. Oligonucleotide sequences used for PCR amplification are listed in Supplementary Table 1. The PCR products from positive clones were ligated to pGEM-T Easy Vector (Promega) and further confirmed by DNA sequencing.

Cell proliferation and survival assay. Equal numbers of cells were plated in 12-well plates in triplicate. Cells were fixed with 10% methanol and stained with 0.1% crystal violet (dissolved in 10% methanol) at indicated times. After staining, wells were washed three times with PBS and destained with acetic acid. The absorbance of the crystal violet solution was measured at 590 nm. For cell survival assay, cells were seeded at a concentration of 1,000 cells per well in 96-well plates and treated with indicated concentrations of α-amanitin or actinomycin D 24 h later. Cell viability was quantified using WST-1 reagent (Roche) according to the manufacturer's instructions. All experiments were performed in triplicate.

Apoptosis and cell cycle analysis. The HCT116 and SNU283 cell lines were treated with or without α-amanitin for 2 days or doxycycline for 4 days at indicated concentrations and stained with annexin-V–phycoerythrin (PE) and 7-AAD (Biovision). Apoptosis was analysed by flow cytometry using a Guava EasyCyte Flow Cytometer (Millipore) according to the manufacturer's protocol. Both pre-apoptotic (annexin-V-positive and 7-AAD-negative) and apoptotic (annexin-V-positive and 7-AAD-positive) cells were included in the analyses. For cell cycle analysis, cells were fixed in 75% ethanol at -20 °C overnight. The cells were washed with cold PBS, treated with 100 μg of RNase A (Qiagen), and stained with 50 μg of propidium iodide (Roche). Cell cycle profiles were analysed by flow cytometry using the Guava EasyCyte Flow Cytometer (Millipore).

Fluorescence in situ hybridization. FISH analysis was performed using fluorescein-labelled POLR2A (red) and control centromere (chromosome 17, green) probes from Empire Genomics. Hybridization and detection were performed according to the manufacturer's protocols. The slides were counterstained with 4′,6-diamidino-2-phenylindole (DAPI), and the images were captured using a Nikon E800 microscope equipped with a cooled-charge coupled device (CCD) camera. To determine hemizygous loss of the POLR2A gene, 100 individual nuclei were analysed for each case. The interphase nuclei were captured and processed using the Quantitative Image Processing System (Applied Imaging).

Patient samples. Matched normal and colorectal tumour tissue samples from patients were obtained from the MD Anderson Cancer Center (MDACC) through appropriate informed consents after approval by the Institutional Review Board (IRB PA11-0767). To determine the expression levels of POLR2A protein, tissue samples (20–40 mg) were placed in tubes containing ceramic beads and were homogenized using a Precellys 24 homogenizer device (Bertin Technologies). The lysates were spun-down twice (15 min, 16,000g) and the supernatant was collected.

Genomic DNA isolation and copy number validation. Total genomic DNA was extracted from human tissue specimens and cell lines using DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's purification instructions. All the DNA samples were stored at -20 °C. The copy number variations for POLR2A were determined using TaqMan probes (Hs02023849_cn and Hs01252684_cn) and standard TaqMan PCR kit on an Applied Biosystems 7900HT Sequence Detection System. And the reference gene TERT was simultaneously quantified in the same reaction for each DNA sample.

Conjugation of α-amanitin to anti-EpCAM antibody (HEA125). Antibody–drug conjugate ama–HEA125 was constructed by coupling of α-amanitin to lysine residues of the HEA125 antibody by a stable linker structure. HEA125 binds to EpCAM-expressing cells with high affinity ($K_d \sim 2.2 \times 10^{-9}$ mol l^{-1}) and high specificity. HEA125 was purified by affinity chromatography using a protein A-Sepharose CL-4B column (GE Healthcare). α-amanitin was attached to immunoglobulin molecules by a plasma stable urea linkage intended to release free α-amanitin inside the tumour cell after lysosomal degradation of the antibody moiety. The drug–antibody ratio of the α-amanitin:IgG molecule was 4:1. Biochemical characteristics of ama–HEA125 were evaluated by high performance liquid chromatography (HPLC) using a PlatinBlue HPLC system (Knauer). In

addition, HEA125 and ama–HEA125 were analysed by reducing SDS–PAGE and Coomassie staining according to common procedures. Verification of drug-loading was done by anti-amanitin immunoblotting analysis of 30 ng HEA125 and ama–HEA125 using standard techniques.

Liposomal nanoparticle preparation. siRNAs for *in vivo* delivery were encapsulated into DOPC. DOPC and siRNA were mixed in the presence of excess tertiary butanol at a ratio of 1:10 (w/w) siRNA:DOPC. Tween 20 was added to the mixture in a ratio of 1:19 Tween 20:siRNA–DOPC. The mixture was vortexed, frozen in an acetone/dry-ice bath, and lyophilized. Before *in vivo* administration, this preparation was hydrated with PBS at room temperature at a concentration of 150–1,000 μg siRNA per kilogram per injection (each mouse received 200 μl of DOPC–siRNA–PBS solution by the intraperitoneal route).

Xenograft tumour studies. Four- to six-week-old female NOD/SCID mice were purchased from Jackson Laboratories and housed under pathogen-free conditions. All studies were approved and supervised by the Institutional Animal Care and Use Committee of MD Anderson Cancer Center. When used in a power calculation, our sample size predetermination experiments indicated that five mice per group can identify the expected effect of POLR2A on tumour size and weight ($P < 0.05$) with 90% power. Animals were randomly divided to different groups. Dox-inducible HCT116 (1×10^6) and SNU283 (2×10^6) cells in 50 μl growth medium (mixed with Matrigel at 1:1) were injected subcutaneously into the flank using a 100-μl Hamilton microlitre syringe. Tumour size was measured every 3 days using a calliper, and tumour volume was calculated using the standard formula: $0.5 \times L \times W^2$, in which L is the longest diameter and W is the shortest diameter. For orthotopic mouse model, the NOD/SCID mice were anaesthetized and the skin was incised to expose caecum. Dox-inducible HCT116 cells (1×10^6) expressing luciferase were injected into the caecal wall using a 100-μl Hamilton microlitre syringe, and then the incision was closed using wound clips. Tumours were monitored by the IVIS system after luciferin injection for 15 min. After initial establishment of tumour (100 mm³ for subcutaneous implants and 2×10^8 photons per second, total flux for orthotopic implants), mice were treated with 1 μg ml⁻¹ doxycycline in drinking water for 3–4 weeks. The doxycycline water was changed every other day.

For xenograft tumour studies using DOPC-encapsulated siRNAs, isogenic pairs of HCT116 (1×10^6) cells were transplanted into the caecal wall using a 100-μl Hamilton microlitre syringe. Ten days after cell injection, mice were randomly divided and assigned to receive either control-siRNA–DOPC or *POLR2A*-siRNA–DOPC. The siRNA sequences are as follows: control siRNA (5′-UUCUCCGAAC GUGUCACGU-3′ and 5′-ACGUGACACGUUCGGAGAA-3′); *POLR2A* siRNA-1 (5′-CCAACAUGCUGACAGAUAU-3′ and 5′-AUAUCUGUCAGCA UGUUGG-3′); *POLR2A* siRNA-2 (5′-CCAAGAAGCGGCUCACACA-3′ and 5′-UGUGUGAGCCGCUUCUUGG-3′). A dose of 150–1,000 μg siRNA per kilogram mouse was administrated intraperitoneally at twice weekly intervals. This range of concentrations was determined by the Center for RNA Interference and Noncoding RNAs at MD Anderson Cancer Center, which ensures efficient delivery and knockdown of target genes as previously described[23].

For xenograft tumour studies using ama–HEA125 antibody–drug conjugate, isogenic pairs of HCT116 (1×10^6), xhCRC (0.5×10^6), SW480 (1×10^6) or SW837 (2×10^6) cells were transplanted into the caecal wall using a 100-μl Hamilton microlitre syringe. Mice bearing 10-day-old tumours were randomized to five groups ($n = 10$ mice per group) and received two intraperitoneal doses (at 10 days and 17 days, 1 week apart) of the following: (1) control unconjugated HEA125 monoclonal antibody at a dose of 3.6 mg kg⁻¹ of body weight; (2) ama–HEA125 at a dose of 90 μg kg⁻¹ in terms of α-amanitin (corresponding to 3.6 mg kg⁻¹ of IgG); (3) ama–HEA125, 30 μg kg⁻¹ in terms of α-amanitin (corresponding to 1.2 mg kg⁻¹ of IgG); (4) ama–HE125, 10 μg kg⁻¹ in terms of α-amanitin (corresponding to 0.4 mg kg⁻¹ of IgG); and (5) ama–HEA125, 3 μg kg⁻¹ in terms of α-amanitin (corresponding to 0.12 mg kg⁻¹ of IgG). Tumours were monitored by the IVIS live imaging system twice a week after luciferin injection for 15 min. Body weights were recorded every four days. Blood was obtained by

retro-orbital bleeding after anaesthesia on day 21 and levels of blood liver enzymes (aspartate amino transferase, amino alanine transferase and alkaline phosphatase) were determined at the Clinical Pathology, Veterinary Medicine and Surgery Core of MD Anderson Cancer Center.

The person who provided animal care and measured tumour growth was blinded to the group allocation during all animal experiments and outcome assessment. Mice were euthanized when they met the institutional euthanasia criteria for tumour size and overall health condition. Tumours were removed, photographed and weighed. The freshly dissected tumour tissues were fixed in 10% buffered formalin overnight, washed with PBS, transferred to 70% ethanol, embedded in paraffin, sectioned and stained with haematoxylin and eosin. The investigators were blinded to the group allocation during the experiment and when assessing the outcome.

Immunohistochemistry and human colon tissue microarray. Colon cancer tissue microarray (BC051110a) was purchased from Biomax, including 110 colon tumour samples and 10 nomral colon tissue samples. Samples were deparaffinized and rehydrated. Antigen was retrieved using 0.01 M sodium citrate buffer (pH 6.0) at a sub-boiling temperature for 10 min after boiling in a microwave oven. To block endogenous peroxidase activity, the sections were incubated with 3% hydrogen peroxide for 10 min. After 1 h of pre-incubation in 5% normal goat serum to prevent nonspecific staining, the samples were incubated with antibody against POLR2A (sc-47701, Santa Cruz), Ki67, (D3B5, Cell Signaling), or cleaved caspase-3 (5A1E, Cell Signaling) at 4 °C overnight. The sections were incubated with a biotinylated secondary antibody (4Plus Biotinylated anti-mouse or anti-rabbit IgG, BioCARE) and then incubated with avidin–biotin peroxidase complex solution and developed using a DAB (diaminobenzidine) substrate kit (550880, BD Biosciences) according to the manufacturer's protocol. Counterstaining colour was carried out using Harris modified haematoxylin. All immunostained slides were scanned on the Automate Cellular Image System III (ACIS III) for quantification by digital image analysis.

Bioinformatic analysis. We analysed the correlation between gene copy number and the corresponding gene expression using data obtained from CCLE (http://www.broadinstitute.org/ccle) and TCGA (http://www.cbioportal.org/public-portal/) as previously described[13]. Enrichment of Pearson correlation coefficients was determined by permuting gene names. To determine whether a deleted gene functions as a housekeeping gene, we first analysed its expression profiles in tumour and normal tissues as well as its general functions from literature. Second, we checked gene conservation across species and lethality of the gene knockout by searching available databases of model organisms (Saccharomyces Genome Database, WormBase, FlyBase and Mouse Genome Informatics). Third, we checked the proximity of the potential target gene to the *TP53* gene and analysed its co-deletion with *TP53* in human cancers. Finally, we searched for available cancer cell lines with the deletion of *TP53* and the target gene to test our hypothesis.

Statistical analysis. Each experiment was repeated three times or more. Unless otherwise noted, data are presented as mean and s.d. or s.e.m., and Student's *t*-test (unpaired, two-tailed) was used to compare two groups of independent samples. In an unpaired *t*-test, we assumed equal variance and no samples were excluded from the analysis. Statistical methods used for TCGA data analysis are described above. $P < 0.05$ was considered statistically significant.

25. Liu, Y. *et al.* Kaposi's sarcoma-associated herpesvirus-encoded microRNA miR-K12-11 attenuates transforming growth factor beta signaling through suppression of SMAD5. *J. Virol.* **86,** 1372–1381 (2012).
26. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
27. Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649,** 247–256 (2010).
28. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8,** 2281–2308 (2013).
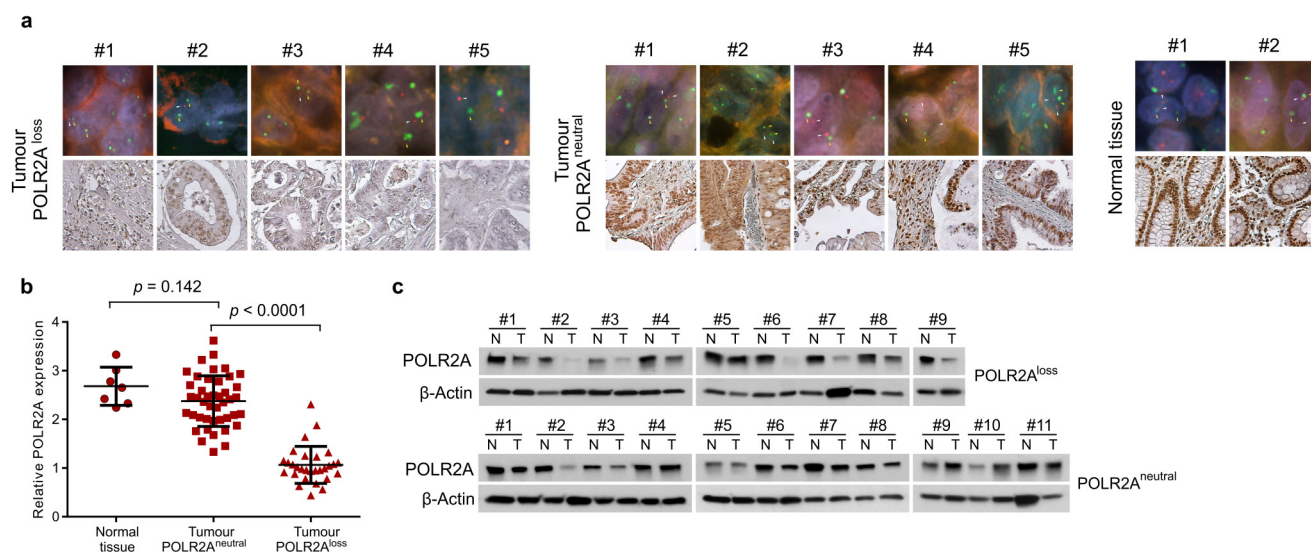
Extended Data Figure 1 | Expression of *POLR2A* correlates with its gene copy number in human colon tumours. a, Top, double-colour FISH analysis using a probe for chromosome 17 centromere (green) and a locus-specific probe for *POLR2A* (red) on human colon tissue microarray. Bottom, immunohistochemical staining of POLR2A in the corresponding tissue samples. Hemizygous loss of the *POLR2A* gene was determined, and the results are shown in Supplementary Table 2. **b**, Quantification of POLR2A expression in human colon normal ($n = 7$), POLR2A^{neutral} ($n = 43$) or POLR2A^{loss} ($n = 29$) tumour tissue samples. Error bars, s.d. **c**, Protein levels of POLR2A and β-actin in matched normal and CRC tissue samples.
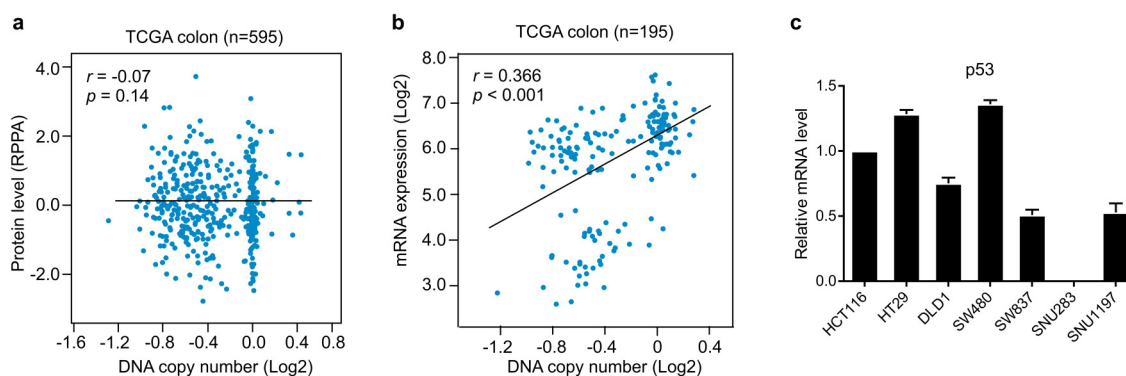
Extended Data Figure 2 | Expression of *TP53* is not associated with its gene copy number. a, **b**, Scatterplots of *TP53* copy number versus protein expression (**a**) or mRNA expression (**b**) in colorectal tumours in TCGA database. Pearson correlation coefficients (*r*) and *P* values are displayed. **c**, Relative mRNA expression of *TP53* in human CRC cell lines (normalized to that in the HCT116 cell line). Data are mean and s.d. of three independent experiments.
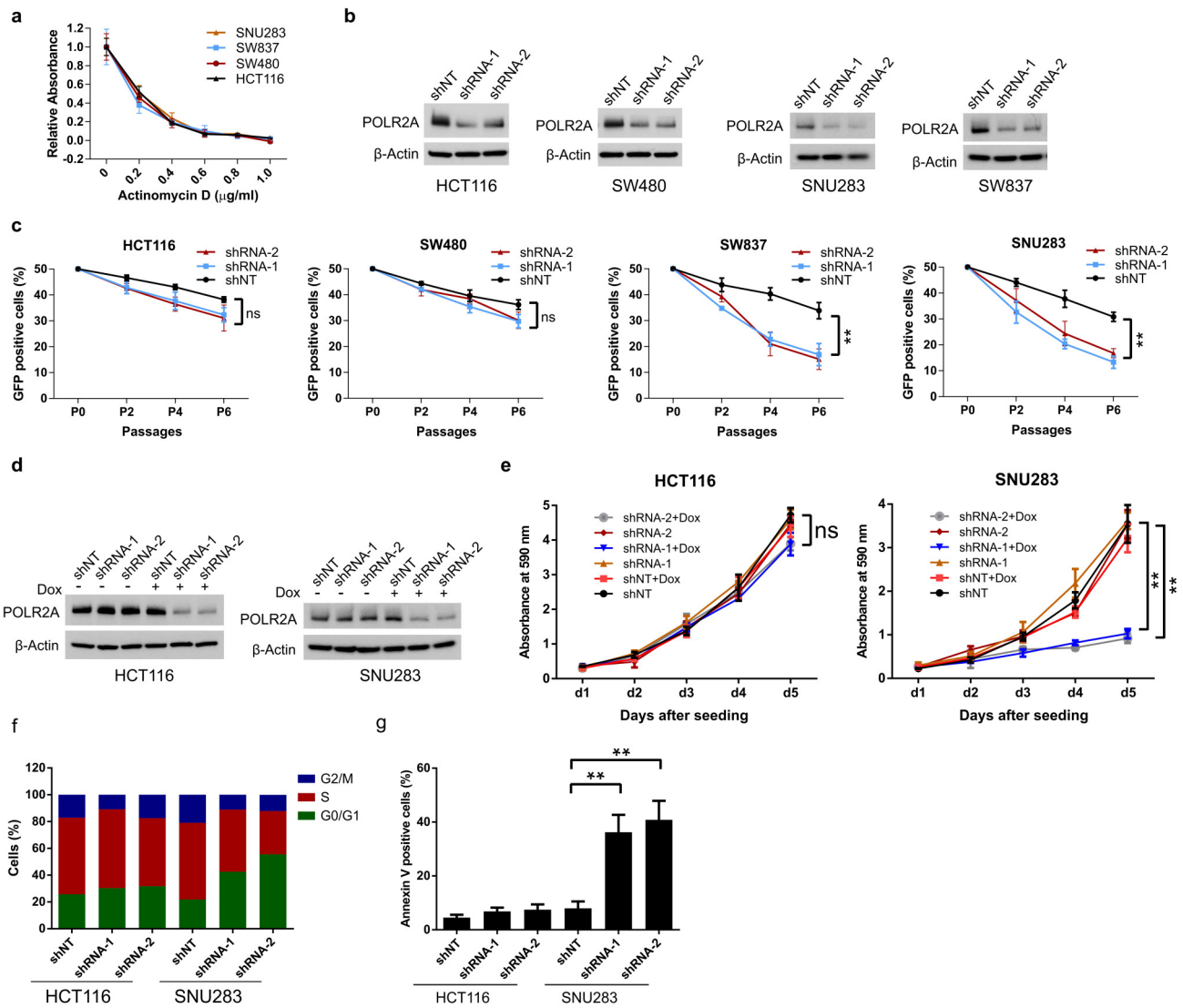
Extended Data Figure 3 | POLR2A^{loss} cells are highly sensitive to POLR2A inhibition. a, Cell proliferation of POLR2A^{neutral} and POLR2A^{loss} cells treated with actinomycin D. **b**, Knockdown efficiency of *POLR2A*-specific shRNAs in HCT116, SW480, SW837 and SNU283 cells. shNT denotes non-targeting control shRNA. **c**, Effect of POLR2A knockdown on the proliferation of four colorectal cancer cell lines. Cells expressing GFP and control or *POLR2A*-specific shRNAs were sorted and mixed with control GFP-negative cells (1:1) and the GFP-positive cells were quantified at passages 2, 4 and 6. **P < 0.01; ns,

not significant. **d**, Protein levels of POLR2A in HCT116 and SNU283 cells expressing Dox-inducible *POLR2A* shRNAs (1.0 μg ml^{-1} Dox). **e**, Cell proliferation of HCT116 and SNU283 cells expressing Dox-inducible *POLR2A* shRNA in the presence of 300 ng ml^{-1} Dox. **P < 0.01. **f**, **g**, Cell cycle profiles (**f**) and apoptosis (**g**) of control or *POLR2A* shRNA-expressing HCT116 and SNU283 cells. **P < 0.01. Data are mean and s.d. of three independent experiments.
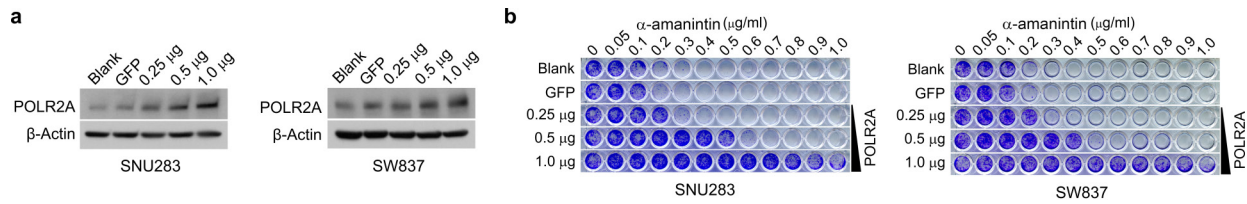
Extended Data Figure 4 | Ectopic expression of POLR2A restores the resistance of POLR2A^loss cells to α-amanitin treatment. **a**, Protein levels of POLR2A in SNU283 and SW837 cells expressing increasing amounts of exogenous POLR2A. **b**, Crystal violet staining of SNU283 and SW837 cells treated with α-amanitin after transfection with increasing amounts of POLR2A expression vector DNA.
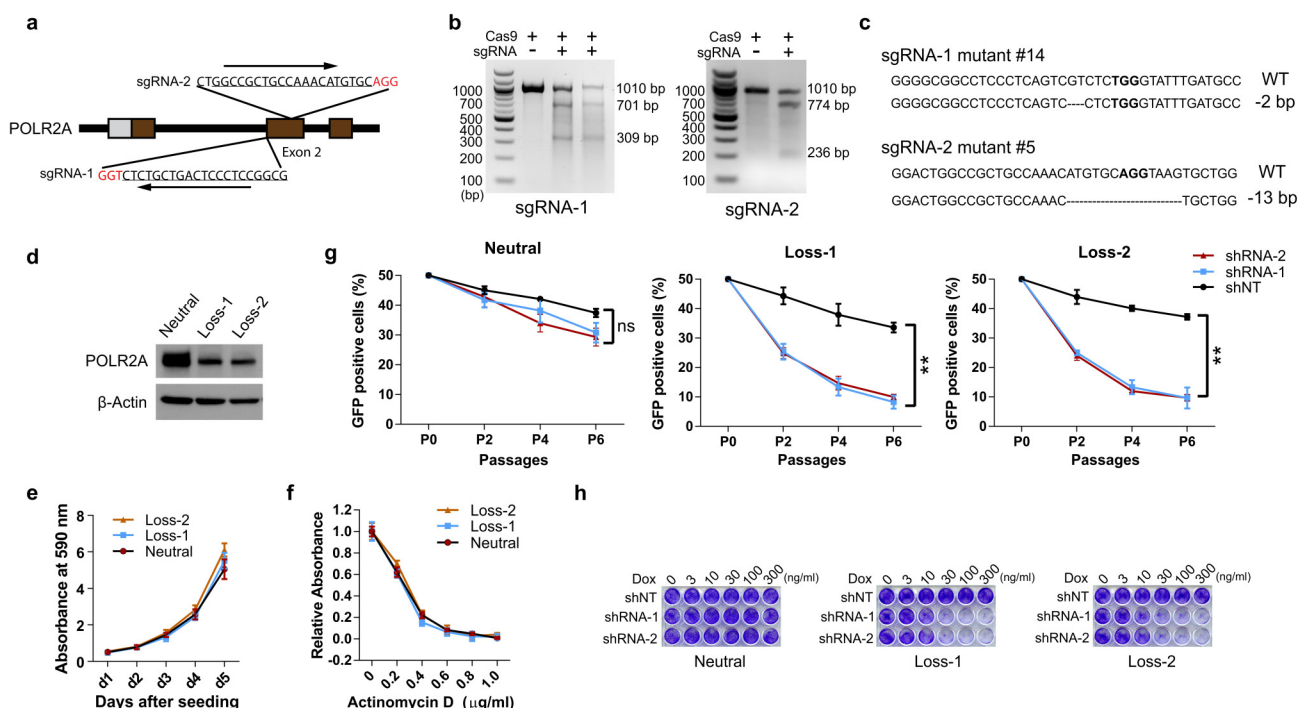
Extended Data Figure 5 | Mono-allelic knockout of POLR2A sensitizes HCT116 cells to POLR2A inhibition. **a**, Schematic illustration of the Cas9/sgRNA-targeting sites in the *POLR2A* gene. Two single-guide RNA (sgRNA)-targeting sequences are shown and the protospacer-adjacent motif (PAM) sequences are highlighted in red. **b**, Efficiency of the Cas9-mediated cleavage of *POLR2A* in HCT116 cells measured by the Surveyor assay. **c**, Sequences of mutant *POLR2A* alleles in the cell colonies 14 and 5. PAM sequences are highlighted in red. Small deletions in the targeted region led to open reading frame shift, producing only a short stretch of the amino-terminal peptide without any functional domains of POLR2A. **d**, Protein levels of POLR2A in POLR2A^{neutral} and POLR2A^{loss} HCT116 cells. **e**, Growth curves of POLR2A^{neutral} and POLR2A^{loss} HCT116 cells. **f**, Relative proliferation of POLR2A^{neutral} and POLR2A^{loss} cells treated with actinomycin D. **g**, Effect of POLR2A knockdown on the POLR2A^{neutral} and POLR2A^{loss} HCT116 cells. Experiments were performed as described in Extended Data Fig. 3c. **$P < 0.01$. **h**, Dox-induced partial suppression of POLR2A inhibited the growth of POLR2A^{loss} HCT116 cells, but not of parental POLR2A^{neutral} HCT116 cells. Data are mean and s.d. of three independent experiments.
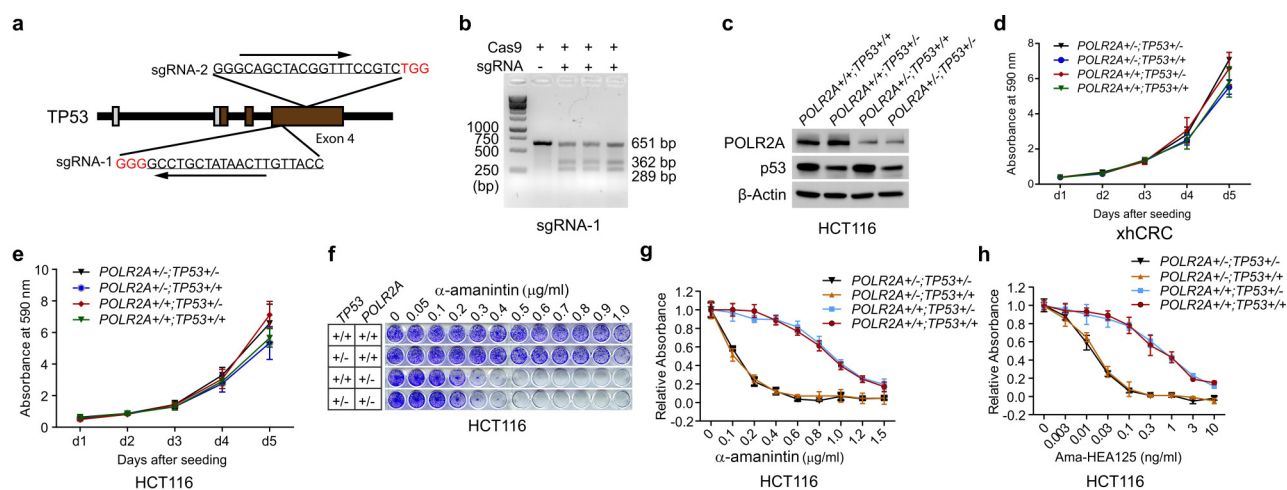
Extended Data Figure 6 | Sensitivity of POLR2A^loss cells to POLR2A inhibition is independent of p53. a, Schematic illustration of the Cas9/sgRNA-targeting sites in the *TP53* gene. Two sgRNA-targeting sequences are shown and the PAM sequences are highlighted in red. **b**, Efficiency of the Cas9-mediated cleavage of *TP53* in HCT116 cells measured by Surveyor assay. **c**, Protein levels of POLR2A and p53 in a panel of isogenic HCT116 cells.

d, Growth curves of POLR2A^neutral and POLR2A^loss xhCRC cells. **e**, Growth curves of POLR2A^neutral and POLR2A^loss HCT116 cells. **f, g**, Crystal staining images (**f**) and cell survival curves (**g**) of POLR2A^neutral and POLR2A^loss HCT116 cells treated with α-amanitin. **h**, Cell survival curves of POLR2A^neutral and POLR2A^loss HCT116 cells in response to the treatment of ama–HEA125. Data are mean and s.d. of three independent experiments.
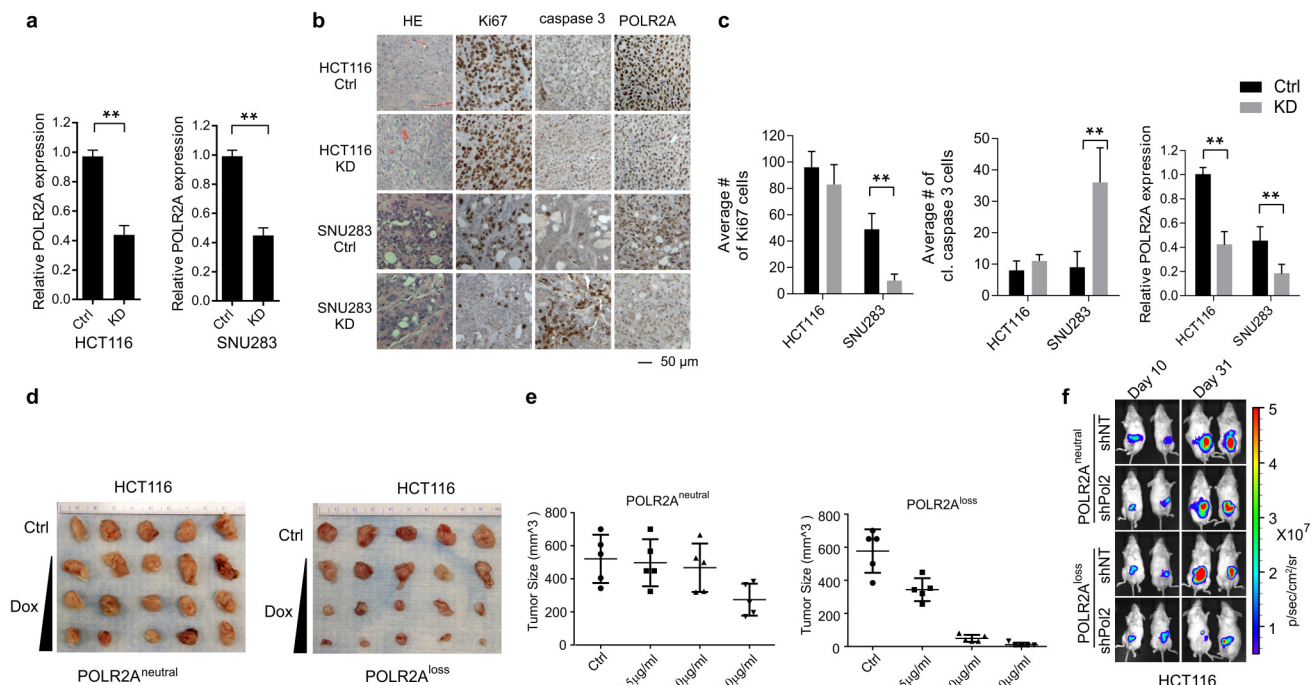
Extended Data Figure 7 | Dose-dependent suppression of POLR2A inhibits tumorigenesis in POLR2A^{loss}, but not POLR2A^{neutral} tumours.
a, Quantification of *POLR2A* mRNA expression levels in subcutaneously xenografted HCT116 and SNU283 tumours expressing control or POLR2A shRNA ($n = 5$ mice per group). **$P < 0.01$. Data are mean and s.d. **b**, Immunohistochemical staining of the aforementioned xenograft tumours. HE, haematoxylin and eosin. **c**, Cells positive for Ki67 (cell proliferation) or cleaved caspase-3 (apoptosis) per field and POLR2A expression in **b** were quantified. **$P < 0.01$. $n = 10$ fields. Data are mean and s.d. **d**, Gross tumour images of xenograft tumours derived from subcutaneously implanted POLR2A^{neutral} and POLR2A^{loss} HCT116 cells (1×10^6 cells injected). Both cell lines express control or Dox-inducible *POLR2A* shRNAs. After the initial establishment of tumours ($100 \, mm^3$), mice were treated with Dox (0.5, 1 and $2 \, \mu g \, ml^{-1}$) in drinking water. $n = 5$ mice per group. **e**, Quantification of tumour sizes as shown in **d**. Data are mean and s.d. **f**, Representative bioluminescent images of orthotopically implanted HCT116 tumours expressing Dox-inducible control or *POLR2A* shRNA after Dox treatment.
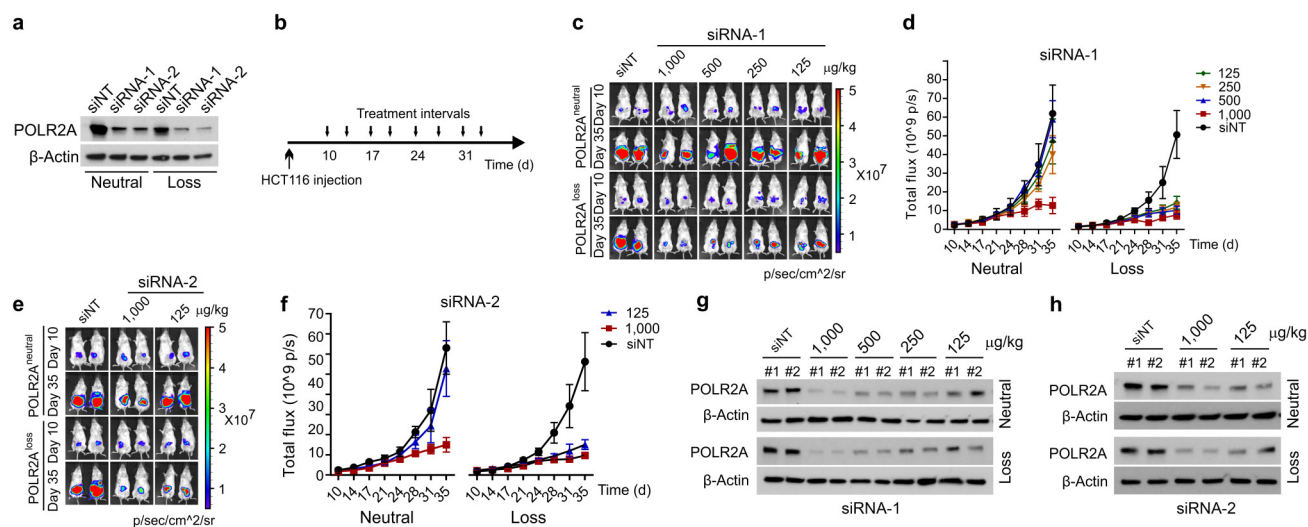
Extended Data Figure 8 | Suppression of POLR2A with DOPC-encapsulated POLR2A siRNA inhibits the growth of POLR2A[loss] tumours.
a, Protein levels of POLR2A after transfection of control siRNA or *POLR2A* siRNAs (shPol2-1 and shPol2-2) in HCT116 cells. **b**, Schematic illustration of orthotopic injection of HCT116 cells (1×10^6 cells) followed by siRNA–DOPC nanoliposome treatment. **c–f**, Representative bioluminescent images (**c**, **e**) and tumour growth curves (**d**, **f**) of orthotopic xenograft tumours derived from POLR2A[neutral] and POLR2A[loss] HCT116 cells that received intraperitoneal injections of control (1,000 µg kg[−1]) or *POLR2A* siRNAs (125, 250, 500 and 1,000 µg kg[−1]) twice weekly. $n = 10$ mice per group. Error bars, s.e.m. **g**, **h**, Representative protein levels of POLR2A in xenograft tumours after control or *POLR2A* siRNA treatment.
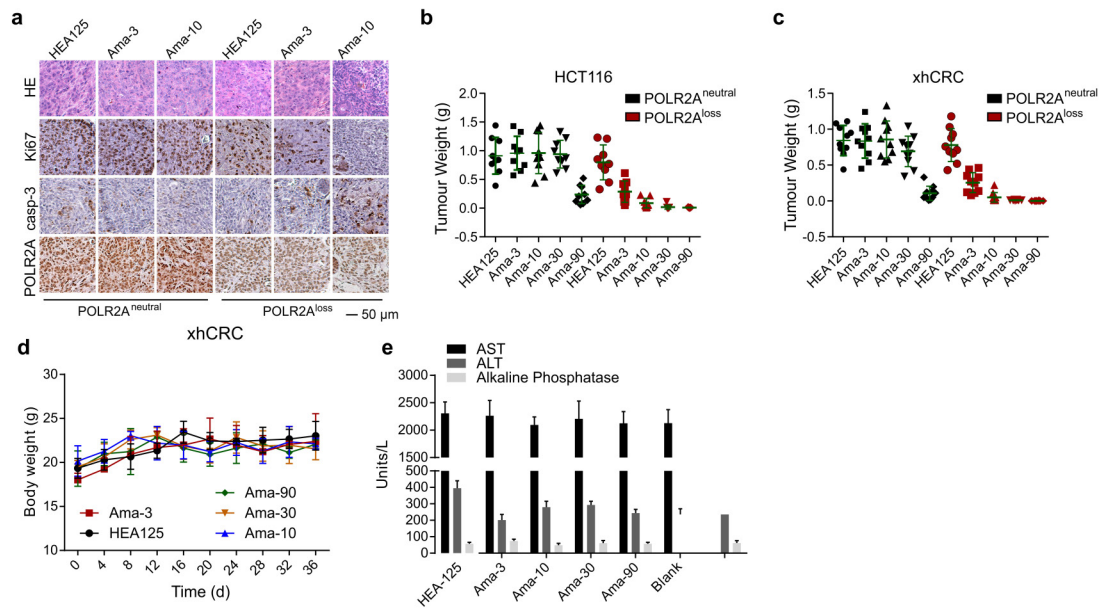
Extended Data Figure 9 | Suppression of POLR2A selectively inhibits the POLR2A^loss tumour growth. a, Immunohistochemical staining of xenografted xhCRC tumours. **b**, **c**, Tumour weights of orthotopically implanted HCT116 (**b**) and xhCRC (**c**) tumours. *n* = 10 mice per group. Data are mean and s.d. **d**, **e**, Body weights (**d**) and liver enzymes (**e**) including alanine aminotransferase (ALT), aspartate aminotransferase (AST) and alkaline phosphatase in peripheral blood. Data are mean and s.d. *n* = 5 mice.
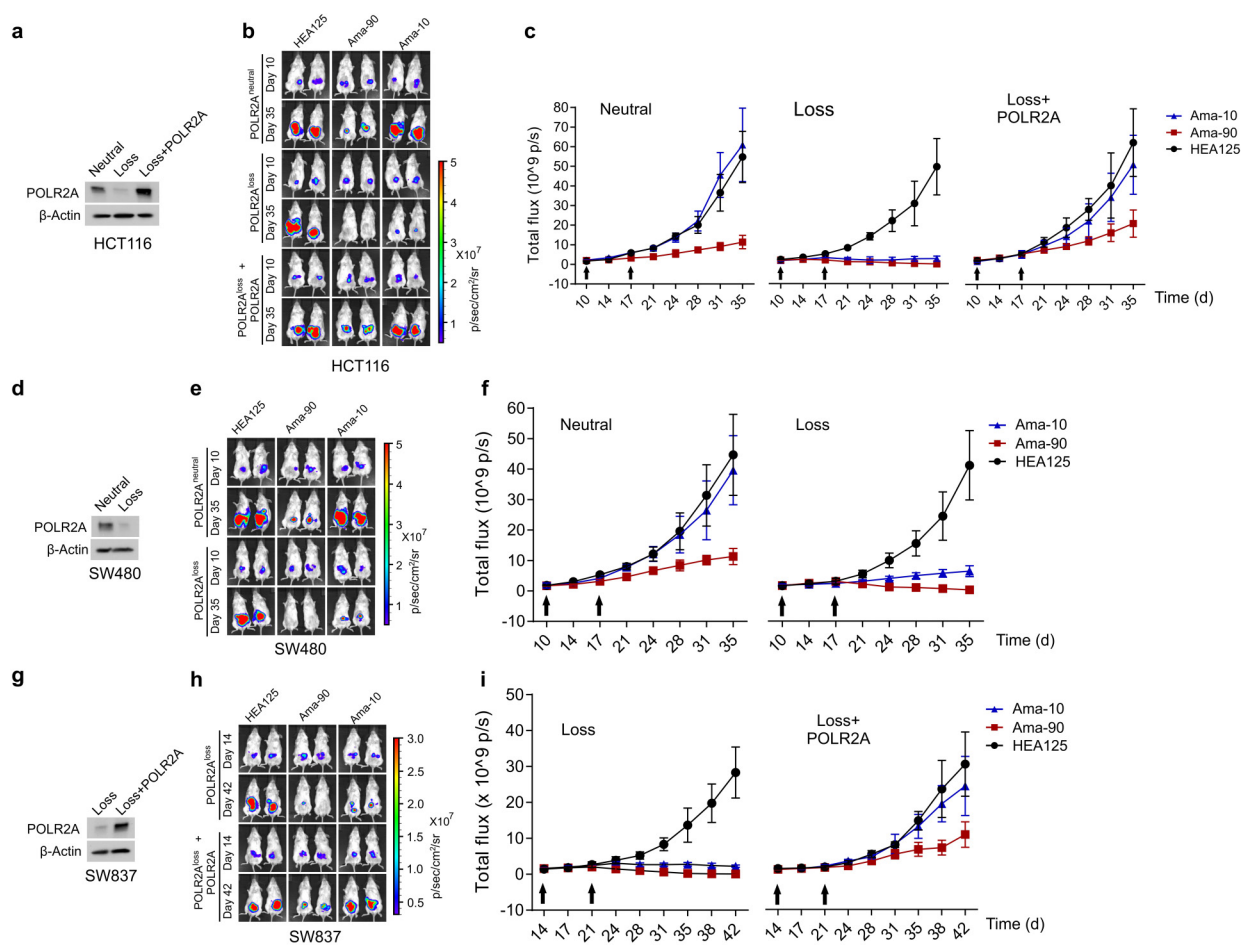
Extended Data Figure 10 | Suppression of POLR2A by ama–HEA125 inhibits the growth of POLR2A^{loss} tumours. a, d, g, Protein levels of POLR2A in HCT116 (**a**), SW480 (**d**) or SW837 (**g**) cells. These cell lines are POLR2A^{neutral}, POLR2A^{loss} or POLR2A-restored. **b, c, e, f, h, i**, Representative bioluminescent images (**b, e, h**) and tumour growth curves (**c, f, i**) of orthotopic xenograft tumours derived from the corresponding cells as indicated. All of them received dual intraperitoneal injections of anti-EpCAM antibody (3.6 mg kg^{-1}) or ama–HEA125 antibody–drug conjugate (10 and 90 µg kg^{-1}, corresponding to 0.4 and 3.6 mg IgG kg^{-1}). $n = 10$ mice per group. Error bars, s.e.m.

LETTER

Structural basis of CpG and inhibitory DNA recognition by Toll–like receptor 9

Umeharu Ohto[1], Takuma Shibata[2,3], Hiromi Tanji[1], Hanako Ishida[1], Elena Krayukhina[4,5], Susumu Uchiyama[4], Kensuke Miyake[2] & Toshiyuki Shimizu[1,3]

Innate immunity serves as the first line of defence against invading pathogens such as bacteria and viruses[1]. Toll-like receptors (TLRs) are examples of innate immune receptors, which sense specific molecular patterns from pathogens and activate immune responses[2]. TLR9 recognizes bacterial and viral DNA containing the cytosine–phosphate–guanine (CpG) dideoxynucleotide motif[3,4]. The molecular basis by which CpG-containing DNA (CpG-DNA) elicits immunostimulatory activity via TLR9 remains to be elucidated. Here we show the crystal structures of three forms of TLR9: unliganded, bound to agonistic CpG-DNA, and bound to inhibitory DNA (iDNA). Agonistic-CpG-DNA-bound TLR9 formed a symmetric TLR9–CpG-DNA complex with 2:2 stoichiometry, whereas iDNA-bound TLR9 was a monomer. CpG-DNA was recognized by both protomers in the dimer, in particular by the amino-terminal fragment (LRRNT–LRR10) from one protomer and the carboxy-terminal fragment (LRR20–LRR22) from the other. The iDNA, which formed a stem-loop structure suitable for binding by intramolecular base pairing, bound to the concave surface from LRR2–LRR10. This structure serves as an important basis for improving our understanding of the functional mechanisms of TLR9.

The immunostimulatory activity of CpG-DNA is affected by its length, the number of CpG motifs, and the sequences flanking the CpG motif. The core CpG motif, which consists of a hexamer with a central unmethylated CpG, has the general formula RRCGYY (where R represents a purine and Y a pyrimidine)[5]. Unlike other TLRs, TLRs 7–9 each have a long insertion loop (Z-loop) between LRR14 and LRR15 (Extended Data Fig. 1). Several recent studies have shown that proteolytic processing at the Z-loop is necessary for the creation of functional TLR9[6–10]. To reveal the molecular mechanism by which TLR9 specifically recognizes CpG-DNA and sends signals to the intracellular compartment, we performed biochemical, biophysical and crystallographic studies of the agonistic- and antagonistic-DNA-binding modes.

We screened recombinantly expressed extracellular domains of TLR9 from various species. Of the proteins we examined, the purified yields of the horse (*Equus caballus*; *Ec*), bovine (*Bos Taurus*; *Bt*) and mouse (*Mus musculus*; *Mm*) orthologues (Extended Data Fig. 1) were sufficient for crystallographic study. We used DNA1668_12mer derived from 20-mer single-stranded DNA1668[3] as agonistic DNA (Extended Data Fig. 2a), and iDNA4084[11] and iDNA_super as iDNAs (Fig. 1a). DNA1668 and iDNAs (iDNA4084 and iDNA_super) function as an agonist and antagonist, respectively, for both horse and bovine TLR9 (Extended Data Fig. 2b). Unless otherwise noted, the DNAs used in this study were single-stranded. Both DNA1668_12mer and iDNA4084 were able to bind Z-loop-processed and -unprocessed TLR9, demonstrating that the binding of both types of DNA is independent of Z-loop processing (Fig. 1b, c). Z-loop-unprocessed TLR9 remained mostly monomeric, irrespective of the presence of agonistic DNA1668_12mer, whereas the proportion of Z-loop-processed TLR9 present as a dimer increased

significantly in the presence of DNA1668_12mer (Fig. 1d and Extended Data Fig. 3). These results indicate that the Z-loop is functionally significant in the ligand-dependent oligomerization of TLR9. By contrast, DNA1668_12mer_GC (DNA1668 with the CpG motif swapped for GpC) and iDNA4084 could induce the formation of only small quantities of TLR9–DNA dimers (Fig. 1d).

We determined the crystal structures of the unliganded, agonistic-DNA-bound, and iDNA-bound forms of Z-loop-processed TLR9 molecules (Extended Data Table 1 and Extended Data Fig. 4). In the crystals, the unliganded and iDNA-bound forms of TLR9 are monomeric (Fig. 2a, b), whereas the agonistic-DNA1668_12mer-bound forms of TLR9 are dimeric (Fig. 2c and Extended Data Fig. 5). Similar to other TLRs[12–17], the agonistic-DNA-bound forms form an m-shaped 2:2 complex in which the C-termini of the two TLR9 protomers are positioned in the centre (Fig. 2c and Extended Data Fig. 5d).

The agonistic DNA binds to two equivalent positions in the dimer, and each DNA1668_12mer is recognized by both TLR9 and TLR9* (The second TLR9 within the dimer and its residues are indicated with asterisks) in a bent conformation (Fig. 2c and Extended Data Fig. 6). DNA1668_12mer winds around the N-terminal fragment of TLR9 from the ascending lateral face to the concave face to interact with a region spanning from LRRNT to LRR10 (interface 1). This binding region is consistent with the results of previous reports[18,19]. Simultaneously, DNA1668_12mer interacts with the loop regions from LRR20–22 in the C-terminal fragment of TLR9* (interface 2) (Extended Data Fig. 6b). Thus, agonistic DNA acts as 'molecular glue' to bridge the two TLR9 molecules. This structural feature strongly suggests that only single-stranded DNA can act as an agonist. Accordingly, double-stranded DNAs containing the CpG motif had greatly reduced affinity (Extended Data Fig. 7 and Extended Data Table 2).

In DNA1668_12mer, the G4–T9 sequence corresponding to the consensus hexamer of GACGTT is mainly recognized by TLR9 as opposed to the other part of DNA. The bases of the CpG motif are accommodated in the groove formed by LRRNT, LRR1 and LRR2 in the ascending lateral face of TLR9 (Fig. 3a). The CpG motif and the flanking bases are recognized via interactions with multiple amino acids, as well as via water-mediated hydrogen bonds (Fig. 3a). The C6 moiety in the CpG motif forms direct hydrogen bonds: the cytosine O2 atom with Met106 N and Ser104 Oγ and the cytosine N3 atom with Ser104 Oγ. In addition, the cytosine N4 atom makes water-mediated (W1 and W2) hydrogen bonds with His76, Pro99 and Phe108, and the cytosine ring itself is wedged between Pro105, Phe108 and a neighbouring CpG guanine (G7) (Fig. 3a). Together with Trp47 and Phe49, C6 forms a three-walled cage that accommodates the neighbouring G7. The G7 N2 atom engages in hydrogen bonds with the Trp96 O and O4 atom of T9 at the +2 position, anchoring the guanine ring to the bottom of the CpG binding groove. Because the N2 atom is unique to guanine these interactions define the specificity for guanine in the CpG motif. The thymine ring

[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [2]Division of Innate Immunity, Department of Microbiology and Immunology, Laboratory of Innate Immunity, Center for Experimental Medicine and Systems Biology, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [3]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Saitama 332-0012, Japan. [4]Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan. [5]U-Medico Corporation, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan.
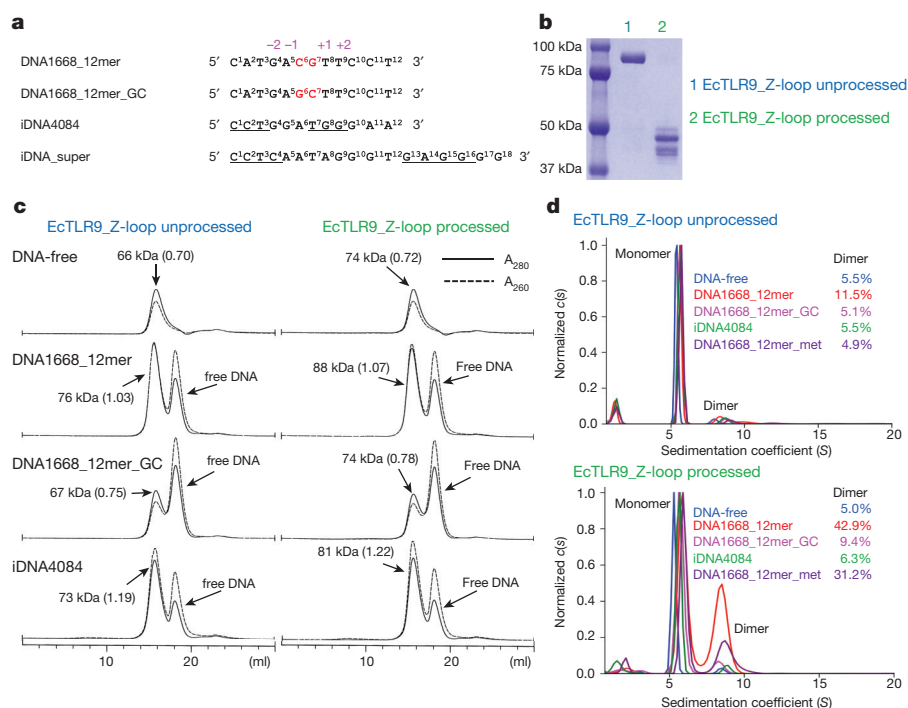
Figure 1 | DNA binding to TLR9 is independent of Z-loop processing, but subsequent oligomerization is dependent on processing. **a**, DNA sequences used in this study. The DNA regions involved in the intramolecular base pairing in iDNA4084 and iDNA_super are underlined. **b**, SDS–PAGE analysis of EcTLR9 with unprocessed and processed Z-loops. **c**, Gel-filtration chromatography of EcTLR9. Black line and dashed line denote absorption at 280 nm and 260 nm, respectively. The ratios of absorbances at 260 nm and 280 nm are shown in parentheses. **d**, The oligomerization states of EcTLR9 analysed by SV–AUC. The original $c(s)$ distributions were normalized against the height of the main peak (see Methods). kDa, kilodalton; S, Svedberg (unit of sedimentation coefficient).

at the +1 position (T8) stacks with Trp47, and the thymine ring at the +2 position (T9) is inserted into the CpG-binding groove and sandwiched between Trp47 and Trp96. T9 also forms hydrogen bonds to Ser72 and G7. The adenine ring at the −1 position (A5) stacks with

Phe108, and in turn G4 (−2 position) stacks onto A5 (Fig. 3a). Because purine bases can form more extensive contacts than pyrimidine bases, the purine–purine sequence preceding the CpG motif is favoured by this three-layered stacking interaction revealed by gel-filtration analysis (Extended Data Fig. 8). The backbone phosphates of G4 and A5 are recognized electrostatically by the positively charged side chains of Lys51, Arg74, His76 and His77. The mutation of several residues (Trp47, Trp96 and Phe108) important for the interaction with DNA resulted in a protein with dramatically reduced binding affinity (Extended Data Table 2), in agreement with the results of the structural analysis.

Interface 2 also plays an important role in the dimerization of TLR9. Residues of interface 2 engage in four hydrogen bonds and several van der Waals interactions (Fig. 3b, c). In contrast to interface 1, which recognizes the base moieties of the CpG motif, LRR20*–LRR22* from the C-terminal fragment primarily recognize the backbone of CpG motif-containing DNA. The insertion loops in the ascending lateral surface of LRR2, LRR5, LRR8, LRR11, LRR18* and LRR20*, which is characteristic of the TLR7–9 family, are involved in protein–protein interactions with complementary shapes (Fig. 3b, d).

We mutated the residues important for the recognition and examined the ability to activate NF-κB (Fig. 3e). Most of the mutants exhibited reduced or completely abolished activation of NF-κB signalling in response to DNA1668. These results clearly demonstrate that interfaces 1 and 2 are both important for the functional integrity of TLR9. We converted the CpG motif into GC, UG, TG and CA and employed isothermal titration calorimetry (ITC) to determine the affinity of TLR9
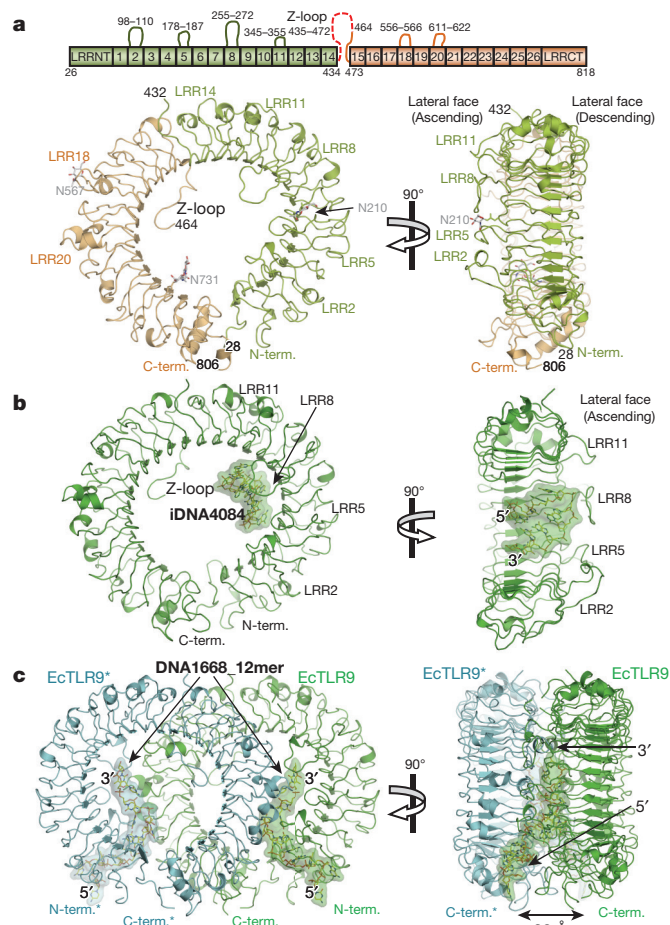
Figure 2 | Structures of TLR9. **a**, Monomer structure of unliganded EcTLR9 showing the lateral face (left) and the convex face from the N-terminal side (right). The *N*-glycan residues are shown in stick representations with their O, N and C atoms coloured red, blue and grey, respectively. The N- and C-terminal halves of EcTLR9 are coloured light green and orange, respectively. **b**, Structure of EcTLR9 in complex with iDNA4084 showing the lateral face (left) and the concave face of the N-terminal half (right). Bound iDNA4084 is shown in stick representation and semi-transparent surface representation. The 5′ and 3′ ends of the DNA are indicated. **c**, Dimer structure of the EcTLR9–DNA1668_12mer complex. EcTLR9 and its dimerization partner EcTLR9* are shown in green and cyan, respectively. The second TLR9 within the dimer and its residues are indicated with asterisks. The DNA1668_12mer molecules are shown in stick representation and semi-transparent surface representation.
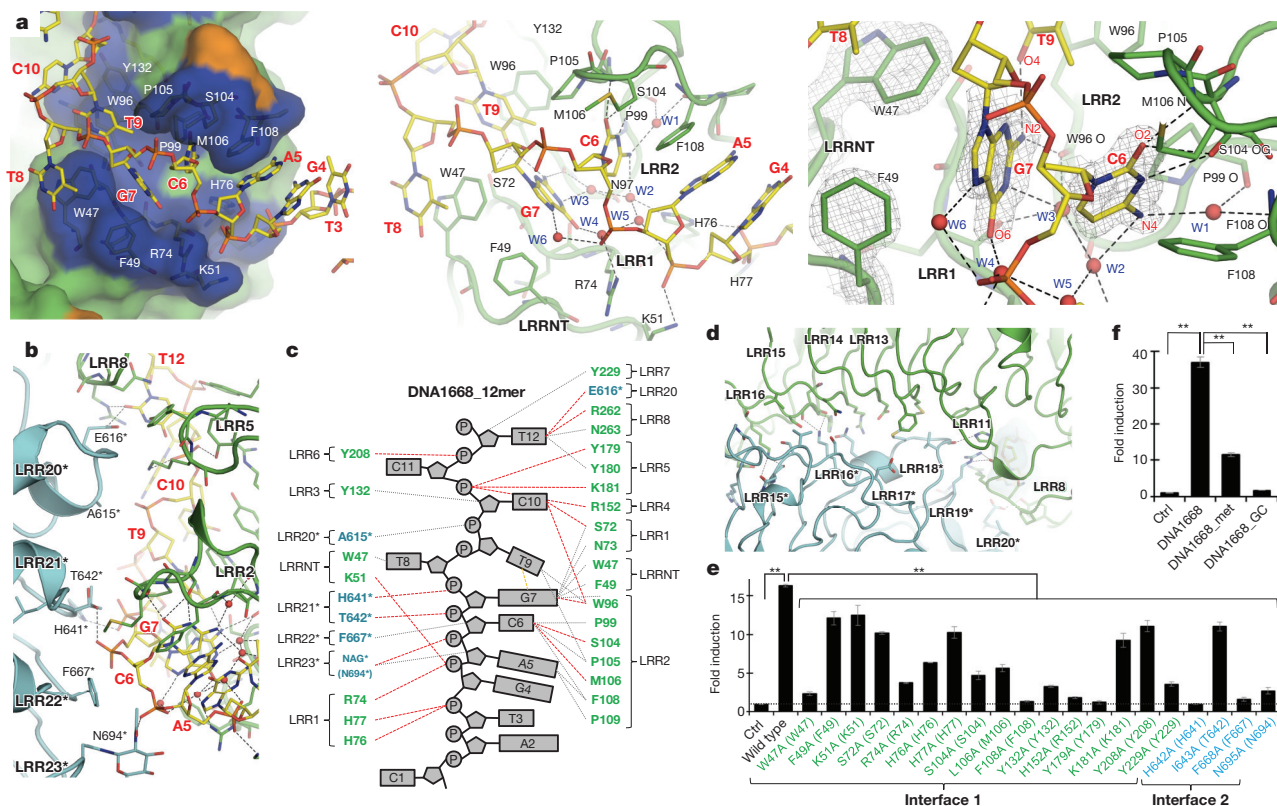
Figure 3 | Agonistic-CpG-DNA recognition by TLR9. a, Left, CpG-DNA binding groove formed by LRRNT, LRR1 and LRR2. Middle, overview of CpG-DNA recognition by TLR9 (EcTLR9_DNA1668_12mer interface 1). Right, magnified view of the CpG recognition. The simulated omit electron densities are shown at a contour level of 3.0σ. **b**, Magnified view of EcTLR9_DNA1668_12mer in interface 2. **c**, Schematic summary of CpG-DNA recognition. Dashed lines in red and black depict hydrogen bonds and van der Waals interactions, respectively. **d**, Magnified view of the protein–protein interface. **e**, NF-κB activation of mouse TLR9 mutants stimulated by agonistic DNA1668 (1 μM). Residues in parenthesis are derived from horse TLR9. 'Ctrl' represents the empty vector without mouse TLR9 cDNA. **f**, The NF-κB activation of mouse TLR9 by DNA1668, DNA1668_GC or DNA1668_met in HEK293T cells. Ctrl, no stimulation. In **e** and **f**, data from triplicate experiments are shown as mean ± s.d. and two-tailed Student's t-test was used to evaluate the statistical significance (**$P < 0.01$).

for the resultant DNAs. The dissociation constant (K_d) values were 20 nM (CG), 569 nM (GC), 54 nM (UG), 163 nM (TG) and 883 nM (CA), respectively (Extended Data Table 2), demonstrating that the CG sequence is important for binding. The mutations of the CpG motif reduced affinities for TLR9: several interactions were disrupted by swapping CG for GC, whereas the direct interactions between G7 (N2) and Trp96 (O) and T9 (O4) were lost by the conversion of CG into CA. C6 (N4) forms a hydrogen bond with the water molecule (W1) that makes hydrogen bonds with Pro99 O and Phe108 O. The substitution of UG for CG would be unfavourable because W1 is surrounded by three hydrogen bond acceptors. The conversion of CG into TG would result in a weak affinity for the same reason as UG, and the methyl group might further weaken an affinity, possibly by disrupting water molecule clusters; the importance of water-mediated interactions for the recognition of methylated DNA in MeCP2 having been previously identified[20]. We also employed ITC to assess the pH-dependence of CpG-DNA binding to TLR9. Binding affinity decreased as pH increased, with K_d values ranging from 20 nM at pH 6.0 to 2500 nM at pH 8.0, revealing that the interaction was stronger under acidic than basic conditions (Extended Data Table 2). Consistent with this, the structural study revealed that His residues are concentrated around the DNA-binding region, resulting in a higher-affinity interaction under acidic conditions (Extended Data Fig. 6c). In addition, we examined the binding affinity of methylated CG (DNA1668_12mer_met) for TLR9. The K_d value for this interaction was 50 nM (Extended Data Table 2), demonstrating that methylated CG yields weaker binding. Accordingly, sedimentation velocity analytical ultracentrifugation (SV–AUC) analysis demonstrated that methylation of the CpG motif exhibited a reduced ability to dimerize (Fig. 1d). Consistent with this, DNA1668_met exhibited reduced activation (Fig. 3f).

We also determined the crystal structures of TLR9 bound to iDNAs (iDNA4084 and iDNA_super) (Fig. 2b and Extended Data Table 1).
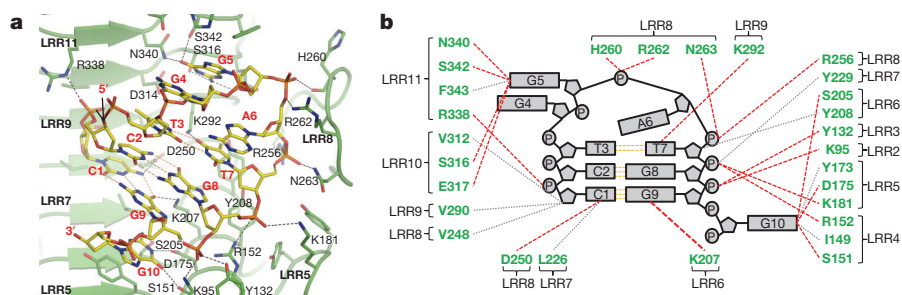
Figure 4 | Recognition of iDNA by TLR9.
a, Magnified view of iDNA4084 bound to EcTLR9. **b**, Schematic summary of iDNA recognition. Dashed lines in red and black depict hydrogen bond and van der Waals interactions, respectively.

The binding site for iDNA partially overlaps with the binding site for agonistic DNA (Extended Data Fig. 6a, b), and this overlap between binding sites accounts for the antagonistic effect of iDNA. Of particular interest, both iDNA4084 and iDNA_super in complex with TLR9, interacting with LRR2–LRR11, form stem-loop structures that fit snugly into the interior of the ring structure of TLR9 (Fig. 4). The stem-loop structure of the iDNAs is formed by intramolecular base pairing between C1-C2-T3 and T7-G8-G9 (two GC pairs and one TT mismatch pair) in the iDNA4084 complex, and between C1-C2-T3-C4 and G13-A14-G15-G16 (three GC pairs and one AT pair) in the iDNA_super complex (Fig. 4 and Extended Data Fig. 4c–e). The length of the loop seems to be immaterial: iDNA_super, which has a long loop, binds similarly to iDNA4084, which has a short loop. The recognition of iDNAs is primarily mediated via the DNA backbone. Also, the base at the cohesive-end position (G10 in iDNA4084) was directly recognized (Fig. 4 and Extended Data Fig. 5f).

The structures of the unliganded, agonistic CpG-DNA-bound and iDNA-bound forms of TLR9 described in this study reveal the structural basis of CpG-DNA recognition and signalling by TLR9, as well as the inhibitory mechanism of iDNA. These results will contribute to the development of therapeutic agents that target TLR9.

1. Janeway, C. A., Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20,** 197–216 (2002).
2. Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Rev. Immunol.* **4,** 499–511 (2004).
3. Bauer, S. *et al.* Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc. Natl Acad. Sci. USA* **98,** 9237–9242 (2001).
4. Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408,** 740–745 (2000).
5. Krieg, A. M. *et al.* CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature* **374,** 546–549 (1995).
6. Ewald, S. E. *et al.* Nucleic acid recognition by Toll-like receptors is coupled to stepwise processing by cathepsins and asparagine endopeptidase. *J. Exp. Med.* **208,** 643–651 (2011).
7. Sepulveda, F. E. *et al.* Critical role for asparagine endopeptidase in endocytic Toll-like receptor signaling in dendritic cells. *Immunity* **31,** 737–748 (2009).
8. Park, B. *et al.* Proteolytic cleavage in an endolysosomal compartment is required for activation of Toll-like receptor 9. *Nature Immunol.* **9,** 1407–1414 (2008).
9. Ewald, S. E. *et al.* The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. *Nature* **456,** 658–662 (2008).
10. Onji, M. *et al.* An essential role for the N-terminal fragment of Toll-like receptor 9 in DNA sensing. *Nature Commun.* **4,** 1949 (2013).
11. Lenert, P. S. Classification, mechanisms of action, and therapeutic applications of inhibitory oligonucleotides for Toll-like receptors (TLR) 7 and 9. *Mediators Inflamm.* **2010,** 986596 (2010).
12. Jin, M. S. *et al.* Crystal structure of the TLR1–TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130,** 1071–1082 (2007).
13. Kang, J. Y. *et al.* Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* **31,** 873–884 (2009).
14. Liu, L. *et al.* Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* **320,** 379–381 (2008).
15. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458,** 1191–1195 (2009).
16. Tanji, H., Ohto, U., Shibata, T., Miyake, K. & Shimizu, T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science* **339,** 1426–1429 (2013).
17. Yoon, S. I. *et al.* Structural basis of TLR5-flagellin recognition and signaling. *Science* **335,** 859–864 (2012).
18. Kubarenko, A. V. *et al.* A naturally occurring variant in human TLR9, P99L, is associated with loss of CpG oligonucleotide responsiveness. *J. Biol. Chem.* **285,** 36486–36494 (2010).
19. Peter, M. E., Kubarenko, A. V., Weber, A. N. & Dalpke, A. H. Identification of an N-terminal recognition site in TLR9 that contributes to CpG-DNA-mediated receptor activation. *J. Immunol.* **182,** 7690–7697 (2009).
20. Ho, K. L. *et al.* MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* **29,** 525–531 (2008).

METHODS

Protein expression, purification and crystallization. The DNA encoding the extracellular domain of Toll-like receptor 9 (TLR9) from various species (human (Q9NR96, residues 25–818, 100%), monkey (F6UZJ0, residues 26–817, 95.7%), horse (Q2EEY0, residues 26–817, 83.6%), bovine (Q866B2, residues 25–815, 77.9%), pig (S5R6V0, residues 25–816, 80.4%), rat (M0RAA8, residues 26–818, 71.5%), mouse (Q9EQU3, residues 26–818, 73.4%) and zebrafish (B3DJW3, residues 23–844, 38.1%)), where values in parentheses corresponds to the Uniprot accession number, the region and sequence identity (versus human) of the extracellular domain, with a C-terminal thrombin cleavage site followed by protein A tag were inserted into the expression vector pMT/BiP/V5-His of the *Drosophila* Expression System. For the preparation of crystallization samples of MmTLR9, a total of seven mutations (N200Q, N242Q, N309Q, N495Q, N568Q, N695Q and N752Q) were introduced to produce the protein with reduced glycosylation sites. *Drosophila* S2 cells were co-transfected with the TLR9 and pCoHygro vectors. Stably transfected cells were selected in Sf-900 II SFM medium containing 300 μg ml^{-1} hygromycin. Z-loop processing of TLR9 is important for its function[6–10]. Therefore, to mimic the Z-loop processing that occurs in the cell, the purification protocol included V8-protease treatment, which yielded the Z-loop-cleaved product. After proteolytic processing, the N- and C-terminal halves of TLR9 remained associated in subsequent purification steps. Protein secreted to the supernatant was captured by IgG Sepharose 6 Fast Flow (GE healthcare) equilibrated with phosphate buffered saline (PBS), washed with ten column volumes of PBS, and eluted by 0.1 M glycine-HCl pH 3.5 and 0.15 M NaCl. Eluent was immediately neutralized by adding with 1/20 volume of 1 M Tris-HCl pH 8.0 and was concentrated to 5–10 mg ml^{-1} and further purified by Superdex 200 gel filtration chromatography equilibrated with 10 mM Tris-HCl pH 7.5 and 0.15 M NaCl. For BtTLR9 and EcTLR9, concentrated TLR9 was added with 1/10 volume of 1 M Na-acetate pH 5.0 and incubated overnight at room temperature with 1–2 U per mg of protein endo H$_f$ (New England Biolabs) for saccharide trimming. Monomeric fractions from Superdex 200 were collected and was incubated with 1/20–1/50 (w/w) V8 protease (Wako) for 12 to 48 h to cleave the Z-loop and protein A tag. Z-loop-processed TLR9 was further purified by HiTrap SP (GE healthcare) cation exchange chromatography. The column was equilibrated with 10 mM Mes pH 6.0 and 0.1 M NaCl and the bound protein was eluted by a linear gradient from 0.1 to 0.7 M NaCl.

For the crystallizations, purified TLR9 was concentrated to 4.0–6.8 mg ml^{-1} in 10 mM Tris (pH 7.5), 150 mM NaCl. To prepare the DNA complex of TLR9, the protein solutions were combined with an approximately twofold excess of DNAs (DNA1668_12mer, iDNA4084 and iDNA_super). Crystallization experiments were performed with sitting-drop vapour-diffusion methods at 293 K. The crystallization droplets were made by mixing the equivolume of protein solution and reservoir solution, typically around 0.5–2.0 μl, except in the case of the EcTLR9–DNA1668_12mer complex where the protein solution and reservoir solution were mixed with to a 3:1 ratio. Corresponding to the observed pH dependency of TLR9, the crystals of agonistic forms of TLR9 complexed with DNA1668_12mer were obtained only in acidic conditions (pH 5.8 and pH 5.5 for horse and bovine TLR9, respectively), while the crystal of unliganded and iDNA bound forms of TLR9 were obtained in wide range of pH (4.5–8.0). The crystallization conditions are summarized in Extended Data Table 1.

Data collection and structure determination. Diffraction data sets were collected on beamlines PF-AR NE3A (Ibaraki, Japan) and SPring-8 BL41XU (Hyogo, Japan) under cryogenic conditions at 100 K. Crystals were soaked into cryoprotectant solution summarized in Extended Data Table 1 and then flash-cooled under a cold gas stream. The diffraction data sets were processed using the HKL2000 package[21] or imosflm[22]. The initial phases for the unliganded form of MmTLR9 were determined with the molecular replacement method by using the program Molrep[23] with the coordinates of the human TLR8 structure (PDB ID: 3W3J)[16]. The model was further refined with stepwise cycles of manual model building using the COOT program[24] and restrained refinement using REFMAC[25] or phenix.refine[26] until the R factors converged. The EcTLR9 and BtTLR9 structures were determined by the molecular replacement method using the Molrep[23] program using the refined MmTLR9 structure. Ligand molecules, *N*-glycans and water molecules were modelled into the electron density maps at the latter cycles of the refinement. The quality of the final structure was evaluated with MolProbity[27]. In the structures of EcTLR9 (unliganded), EcTLR9–DNA1668_12mer, EcTLR9–iDNA4084, BtTLR9–DNA1668_12mer, MmTLR9 (unliganded), MmTLR9–iDNA4084 (form1), MmTLR9–iDNA4084 (form2) and MmTLR9–iDNA_super, 100%, 100%, 98%, 99%, 99%, 99%, 100% and 99% of the residues were in Ramachandran favoured or allowed regions, respectively. The statistics of the data collection and refinement are summarized in Extended Data Table 1. The figures representing structures were prepared with PyMOL[28].

Isothermal titration calorimetry. ITC experiments were performed at 298 K in a buffer composed of 50 mM phosphate buffer pH 6.0–8.0, 250 mM NaCl using a MicroCal iTC$_{200}$ (GE Healthcare). DNAs at a concentration of 50 μM were titrated into 5 μM of wild-type or mutant horse TLR9. The titration sequence included a single 0.4 μl injection followed by 19 injections, 2 μl each, with a spacing of 120 s between the injections. OrigineLab software (GE Healthcare) was used to analyse the raw ITC data. Thermodynamic parameters were extracted from curve fitting analysis with a single-site binding model.

Analytical ultracentrifugation sedimentation velocity. SV–AUC analyses were performed in a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter) equipped with a 4-hole An60Ti rotors at 20 °C using Beckman Coulter 12-mm double-sector charcoal-filled epon centerpieces and sapphire windows. The scanning at 42,000 r.p.m. was performed as quickly as possible between 6.0 and 7.2 cm from the axis of rotation with a radial increment of 30 μm. To analyse the dimerization induced by DNA binding, horse TLR9 Z-loop-unprocessed and Z-loop-processed samples were run at a loading concentration of 20 μM with or without equimolar concentrations of DNAs (Fig. 1d). To analyse the concentration dependence of the dimerization, the AUC measurements were performed at protein concentrations ranging from 1.5 μM to 30 μM in the presence of equimolar DNA1668_12mer (Extended Data Fig. 3). All sedimentation velocity experiments were conducted in a buffer containing 10 mM MES and 250 mM NaCl at pH 5.5.

The sedimentation coefficient distributions were obtained using the $c(s)$ method of SEDFIT[29]. The sedimentation coefficients ranging from 0.1 to 50 S with a logarithmically spaced grid and resolution of 500 were used. The frictional ratio, meniscus, radial and time-invariant noise were floated during the fitting procedure, and a regularization level of 0.68 was used. The partial specific volume, the buffer density and viscosity were calculated using the program SEDNTERP 1.09 and were 0.7407 cm^3 g^{-1}, 1.00852 g ml^{-1} and 1.0256 cP, respectively. The percentages of monomer and dimer were calculated by dividing the corresponding peak area by the sum of the areas under two peaks.

NF-κB-dependent luciferase reporter assay. To check mouse TLR9 response, HEK293T cells were seeded in collagen-coated 6-well plates at a density of 5×10^5 cells per well, and transiently transfected with wild-type or mutant mouse TLR9 cDNAs in pMX-puro-IRES-rat CD2 (1 μg), together with wild-type mouse Unc93B1 cDNA in pMX-puro (0.5 μg) and a pELAM1-luc reporter plasmid (5 ng), using PEI (Polyethylenimine "Max", MW40,000 ; Polysciences, Inc.) at 36 h before stimulation. To check horse and bovine TLR9 response, HEK293T cells were plated on collagen-coated 10-cm dishes at a density of 6×10^6 cells per well, and transiently transfected with wild-type horse or bovine TLR9 cDNAs in pMX-puro-IRES-rat CD2 (3 μg), together with wild-type human Unc93B1 cDNA in pMX-puro (6 μg) and a pELAM1-luc reporter plasmid (30 ng), using PEI at 30 and 24 h before stimulation. The NF-κB luciferase reporter plasmid, pELAM1-luc, was provided by T. Muta (University of Tohoku, Japan)[30]. Twenty-four hours after first transfection, cells were reseeded in collagen-coated flat 96-well plates (Corning) at a density of 1×10^5 cells per well. Then, after pre-culture for 4~6 h, attached cells were stimulated with various DNAs or 100 ng ml^{-1} recombinant human TNF-α (Wako Pure Chemical Industries) for 6 h. Stimulated cells were lysed by 40 μl of Cell Culture Lysis Reagent (Promega) and 6 μl of lysate was subjected to a luciferase assay using the Luciferase Assay System (Promega). The relative light unit (RLU) of chemiluminescence was measured by GloMax 96 Microplate Luminometer (Promega). Since retroviral vector pMX-puro-IRES-rat CD2 allowed for indirect validation of inserted cDNA expression by checking rat CD2 expression, transfection efficiency of wild-type and mutant mouse TLR9 cDNAs in HEK293T cells was evaluated by examination of cell-surface rat CD2 expression level using a FACSCalibur flow cytometer (BD Biosciences). The activity of pELAM1-luc in each transfected HEK293T cells was also verified by stimulation of recombinant human TNF-α.

Oligonucleotide. Oligonucleotides for the gel-filtration, ITC, AUC and crystallographic analyses were single-stranded DNAs with normal phosphodiester linkage unless otherwise noted. Oligonucleotides for the luciferase reporter assay were single-stranded DNAs with phosphorothioate linkage. Phosphorothioate DNAs (DNA1668, DNA1668_GC, DNA1668_12mer and DNA1668_12mer_GC) were purchased from FASMAC (Kanagawa, Japan). Other phosphorothioate DNAs (iDNA4084, iDNA_super and DNA1668_met) and all phosphodiester DNAs were purchased from Eurofins MWG Operon (Ebersberg, Germany).

Statistical analysis. Data from triplicate samples in Fig. 3e, 3f and Extended Data Fig. 2 were shown as mean ± s.d. and subjected to statistical analysis. Statistical significance was determined by two-tailed Student's *t*-tests. A *P* value of less than 0.01 was considered to be significant. No statistical method was used to predetermine sample size.

21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
22. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67,** 271–281 (2011).
23. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66,** 22–25 (2010).

24. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
25. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
26. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
27. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
28. DeLano, W. L. The PyMOL Molecular Graphics System. DeLano Scientific LLC. http://www.pymol.org (2008).
29. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).
30. Muta, T. & Takeshige, K. Essential roles of CD14 and lipopolysaccharide-binding protein for activation of toll-like receptor (TLR)2 as well as TLR4 reconstitution of TLR2- and TLR4-activation by distinguishable ligands in LPS preparations. *FEBS J.* **268,** 4580–4589 (2001).
31. Xu, C., Bian, C., Lam, R., Dong, A. & Min, J. The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nat. Commun.* **2,** 227 (2011).
32. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455,** 818–821 (2008).
33. Avvakumov, G. V., Walker, J. R. & Xue, S. Li, Y. Duan, S., Bronner, C., Arrowsmith, C. H. & Dhe-Paganon, S. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455,** 822–825 (2008).
34. Hashimoto, H., Horton, J. R., Zhang, X. & Cheng, X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455,** 826–829 (2008).

```
Signal Sequence
Human_ TLR9    MGFCRSALHPLSLLVQAIMLAMTLA                         25
Horse_ TLR9    MGPCHGALQPLSLLVQAAMLAVALA                         25
Bovine_TLR9    MGP-YCAPHPLSLLVQAAALAAALA                         24
Mouse_ TLR9    MVLRRRTLHPLSLLVQAAVLAETLA                         25

LRRNT
Human_ TLR9    LGTLPAFLPCELQPHGLVNCNWLFLKSVPHFSMAAPRG            63
Horse_ TLR9    QGTLPPFLPCELQPHGLVNCNWLFLKSVPHFSAAAPRD            63
Bovine_TLR9    EGTLPAFLPCELQPHGQVDCNWLFLKSVPHFSAGAPRA            62
Mouse_ TLR9    LGTLPAFLPCELKPHGLVDCNWLFLKSVPRFSAAASCS            63

LRR1          xLxxLxLxxN
Human_ TLR9    NVTSLSLSSNRIHHLHDSDFAHLP                          87
Horse_ TLR9    NVTSLSLLSNRIHHLHDSDFAQLS                          87
Bovine_TLR9    NVTSLSLISNRIHHLHDSDFVHLS                          86
Mouse_ TLR9    NITRLSLISNRIHHLHNSDFVHLS                          87

LRR2          xLxxLxLxxN
Human_ TLR9    SLRHLNLKWNCPPVGLSPMHFPCHMTIEPSTFLAVP             123
Horse_ TLR9    NLQKLNLKWNCPPAGLSPMHFPCHMTIEPNTFLAVP             123
Bovine_TLR9    NLRVLNLKWNCPPAGLSPMHFPCRMTIEPNTFLAVP             122
Mouse_ TLR9    NLRQLNLKWNCPPTGLSPLHFSCHMTIEPRTFLAMR             123

LRR3          xLxxLxLxxN
Human_ TLR9    TLEELNLSYNNIMTVPALPK                             143
Horse_ TLR9    TLEELNLSYNGITTVPALPS                             143
Bovine_TLR9    TLEELNLSYNGITTVPALPS                             142
Mouse_ TLR9    TLEELNLSYNGITTVPRLPS                             142

LRR4          xLxxLxLxxN
Human_ TLR9    SLISISLSHTNILMLDSASLAGLH                         167
Horse_ TLR9    SLVSILSRTNILQLDPTSLTGLH                          167
Bovine_TLR9    SLVSLSLSHTSILVLGPTHFTGLH                         166
Mouse_ TLR9    SLVNLSLSHTNILVLDANSLAGLY                         167

LRR5          xLxxLxLxxN
Human_ TLR9    ALRFLFMDGNCYYKNPCRQALEVAPGALLGLG                 199
Horse_ TLR9    ALRFLYMDSNCYYKNPCGRALEVAPGALLGLG                 199
Bovine_TLR9    ALRFLYMDSNCYYMNPCPRALEVAPGALLGLG                 198
Mouse_ TLR9    SLRVIFMDGNCYYKNPCTGAVKVTPGALLGLS                 199

LRR6          xLxxLxLxxN
Human_ TLR9    NLTHLSLKYNNLTVVPRNLPS                            220
Horse_ TLR9    NLTHLSLKYNNLTTVPRSLPP                            220
Bovine_TLR9    NLTHLSLKYNNLTEVPRRLPP                            219
Mouse_ TLR9    NLTHLSLKYNNLTKVPRQLPP                            220

LRR7          xLxxLxLxxN
Human_ TLR9    SLEYLLLSYNRIVKLAPEDLANLT                         244
Horse_ TLR9    SLEYLLLSYNHIVTLAPEDLANLT                         244
Bovine_TLR9    SLDTLLLSYNHIVTLAPEDLANLT                         243
Mouse_ TLR9    SLEYLLVSYNLIVKLGPEDLANLT                         244

LRR8          xLxxLxLxxN
Human_ TLR9    ALRVLDVGGNCRRCDHAPNPCMECPRHFPQLHPDTFSHLS         284
Horse_ TLR9    ALRVLDVGGNCRRCDHARNPCVECPHKFPQLHSDTFSHLS         284
Bovine_TLR9    ALRVLDVGGNCRRCDHARNPCRECPKNFPKLHPDTFSHLS         283
Mouse_ TLR9    SLRVLDVGGNCRRCDHAPLPCIECGQKSLHLHPETFHHLS         284

LRR9          xLxxLxLxxN
Human_ TLR9    RLEGLVLKDSSLSWLNASWFRGLG                         308
Horse_ TLR9    RLEGLVLKDSSLYQLNPRWFRGLG                         308
Bovine_TLR9    RLEGLVLKDSSLYKLEKDWFRGLG                         307
Mouse_ TLR9    HLEGLVLKDSSLHTLNSSWFQGLV                         308

LRR10         xLxxLxLxxN
Human_ TLR9    NLRVLDLSENFLYKCITKTKAFQGLT                       334
Horse_ TLR9    NLTVLDLSENFLYDCITKTKAFQGLA                       334
Bovine_TLR9    RLQVLDLSENFLYDYITKTTIFNDLT                       333
Mouse_ TLR9    NLSVLDLSENFLYESITHTNAFQNLT                       334

LRR11         xLxxLxLxxN
Human_ TLR9    QLRKLNLSFNYQKRVSFAHLSLAPSFGSLV                   364
Horse_ TLR9    QLRRLNLSFNYHKKVSFAHLTLAPSFGSLL                   364
Bovine_TLR9    QLRRLNLSFNYHKKVSFAHLHLASSFGSLV                   363
Mouse_ TLR9    RLRKLNLSFNYRKKVSFARLHLASSFKNLV                   364

LRR12         xLxxLxLxxN
Human_ TLR9    ALKEILDMHGIFFRSLDETTLRPLARLP                     391
Horse_ TLR9    SLQEIDMHGIFFRSLSQKTLQPLARLP                     391
Bovine_TLR9    SLEKIDMHGIFFRSLTNITLQSLTRLP                     390
Mouse_ TLR9    SLQEINMNGIFFRLLNKYTLRWLADLP                     391

LRR13         xLxxLxLxxN
Human_ TLR9    MLQTLRLQMNFINQAQLGIFRAFP                         415
Horse_ TLR9    MLQRLYLQMNFINQAQLGIFKDFP                         415
Bovine_TLR9    KLQSLHLQLNFINQAQLSIFGAFP                         414
Mouse_ TLR9    KLHTLHLQMNFINQAQLSIFGTFR                         415

LRR14         xLxxLxLxxN
Human_ TLR9    GLRYVDLSDNRISGASEL-T                             434
Horse_ TLR9    GLRYIDLSDNRISGAVEP-V                             434
Bovine_TLR9    SLLFVDLSDNRISGAATP-A                             433
Mouse_ TLR9    ALRFVDLSDNRISGPSTLSE                             435

Z-loop
Human_ TLR9    ATMGEADGGEKVWLQPGDLAPAPVDTPSSEDFRPNCST           472
Horse_ TLR9    ATTGEVDGGKKVWLTSRDLTPGPLDTPSSEDFMPSCKN           472
Bovine_TLR9    AALGEVDSRVEVWRLPRGLAPGPLDAVSSKDFMPSCN-           470
Mouse_ TLR9    ATPEEADDAEQEELLSADPHPAPLSTPASKNFMDRCKN           473
```

```
LRR15         xLxxLxLxxN
Human_ TLR9    LNFTLDLSRNNLVTVQPEMFAQLS                         496
Horse_ TLR9    LSFTLDLSRNNLVTVQPEMFAQLS                         496
Bovine_TLR9    LNFTLDLSRNNLVTIQQEMFTRLS                         494
Mouse_ TLR9    FKFTMDLSRNNLVTIKPEMFVNLS                         497

LRR16         xLxxLxLxxN
Human_ TLR9    HLQCLRLSHNCISQAVNGSQFLPLT                        521
Horse_ TLR9    RLQCLRLSHNSISQAVNGSQFVPLT                        521
Bovine_TLR9    RLQCLRLSHNSISQAVNGSQFVPLT                        519
Mouse_ TLR9    RLQCLSLSHNSIAQAVNGSQFLPLT                        522

LRR17         xLxxLxLxxN
Human_ TLR9    GLQVLDLSHNKLDLYHEHSFTELP                         545
Horse_ TLR9    SLQVLDLSHNKLDLYHGRSFTELP                         545
Bovine_TLR9    SLRVLDLSHNKLDLYHGRSFTELP                         543
Mouse_ TLR9    NLQVLDLSHNKLDLYHWKSFSELP                         546

LRR18         xLxxLxLxxN
Human_ TLR9    RLEALDLSYNSQPFGMQGVGHNFSFVAHLR                   575
Horse_ TLR9    RLEALDLSYNSQPFSMRGVGHNLSFVAQLP                   575
Bovine_TLR9    QLEALDLSYNSQPFSMQGVGHNLSFVAQLP                   573
Mouse_ TLR9    QLQALDLSYNSQPFSMKGIGHNFSFVTHLS                   576

LRR19         xLxxLxLxxN
Human_ TLR9    TLRHLSLAHNNIHSQVSQQLCST                          598
Horse_ TLR9    TLRYLSLAHNGIHSRVSQQLCST                          598
Bovine_TLR9    SLRYLSLAHNGIHSRVSQKLSSA                          596
Mouse_ TLR9    MLQSLSLAHNDIHTRVSSHLNSN                          599

LRR20         xLxxLxLxxN
Human_ TLR9    SLRALDFSGNALGHMWAEGDLYLHFFQGLS                   628
Horse_ TLR9    SLWALDFSGNSLSQMWAEGDLYLRFFQGLR                   628
Bovine_TLR9    SLRALDFSGNSLSQMWAEGDLYLCFFKGLR                   626
Mouse_ TLR9    SVRFLDFSGNGMGRMWDEGGLYLHFFQGLS                   629

LRR21         xLxxLxLxxN
Human_ TLR9    GLIWLDLSQNRLHTLLPQTLRNLPK                        653
Horse_ TLR9    SLIRLDLSQNRLHTLLPCTLGNLPK                        653
Bovine_TLR9    NLVQLDLSENHLHTLLPRHLDNLPK                        651
Mouse_ TLR9    GLLKLDLSQNNLHILRPQNLDNLPK                        654

LRR22         xLxxLxLxxN
Human_ TLR9    SLQVLRLRDNYLAFFKWWSLHFLP                         677
Horse_ TLR9    SLQLRLRNNYLAFFNWSSLTLLP                          677
Bovine_TLR9    SLRQLRLRDNNLAFFNWSSLTVLP                         675
Mouse_ TLR9    SLKLLSLRDNYLSFFNWTSLSFLP                         678

LRR23         xLxxLxLxxN
Human_ TLR9    KLEVLDLAGNQLKALTNGSLPAGT                         701
Horse_ TLR9    NLETLDLAGNQLKALSNGSLPSGT                         701
Bovine_TLR9    RLEALDLAGNQLKALSNGSLPPGI                         699
Mouse_ TLR9    NLEVLDLAGNQLKALTNGTLPNGT                         702

LRR24         xLxxLxLxxN
Human_ TLR9    RLRRLDVSCNSISFVAPGFFSKAK                         725
Horse_ TLR9    QLQRLDVSRNSIIFVVPGFFALAT                         725
Bovine_TLR9    RLQKLDVSSNSIGFVIPGFFVRAT                         723
Mouse_ TLR9    LLQKLDVSSNSIVSVVPAFFALAV                         726

LRR25         xLxxLxLxxN
Human_ TLR9    ELRELNLSANALKTVDHSWFGPLAS                        750
Horse_ TLR9    RLRELNLSANALRTVEPSWFGFLAG                        750
Bovine_TLR9    RLIELNLSANALKTVDPSWFGSLAG                        748
Mouse_ TLR9    ELKEVNLSHNILKTVDRSWFGPIVM                        751

LRR26         xLxxLxLxxN
Human_ TLR9    ALQILDLVSANPLH                                   763
Horse_ TLR9    SLEVLDVSANPLH                                    763
Bovine_TLR9    TLKILDVSANPLH                                    761
Mouse_ TLR9    NLTVLDVRSNPLH                                    764

LRRCT
Human_ TLR9    CACGGAAFMDFLLEVQAAVPGLPSRVKCGSPGQLQGLSIFAQDLRLCLDEALSWDC   818
Horse_ TLR9    CACGGAAFVDFLLQVQAAVPGLPSRVKCGSPGQLQGRSIFAQDLRLCLDESLSWDC   818
Bovine_TLR9    CACGGAAFVDFLLEREAVPGLSRRVTCGSPGQLQGRSIFTQDLRLCLDETLSLDC   816
Mouse_ TLR9    CACGGAAFVDLLLEVQTKVPGLANGVKCGSPGQLQGRSIFAQDLRLCLDEVLSWDC   819
```

Agonist DNA interface1

Agonist DNA interface2

Antagonist DNA interface

Protein-protein interface (▬▬)

Extended Data Figure 1 | Sequence alignment of human, horse, bovine and mouse TLR9. Sequence alignments are displayed for each LRR module. The agonist DNA interfaces 1 and 2 deduced from the EcTLR9–DNA1668_12mer complex are indicated by blue and yellow highlighting, respectively. The antagonist DNA interface deduced from the EcTLR9–iDNA4084 complex is indicated with boxes. The protein–protein interface in the EcTLR9–DNA1668_12mer complex is indicated by bold orange bars below each LRR module. Alignments were performed using Clustal Omega software (EMBL-European Bioinformatics Institute). Residues are coloured to indicate the degree of similarity: red residues are those with the highest similarity, followed by green, blue and black (lowest similarity).
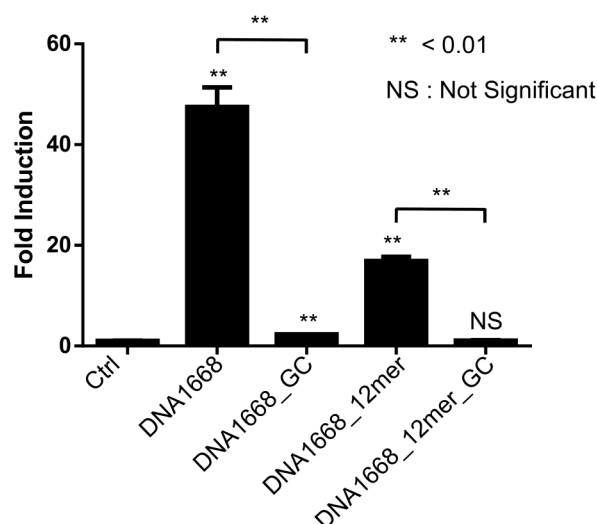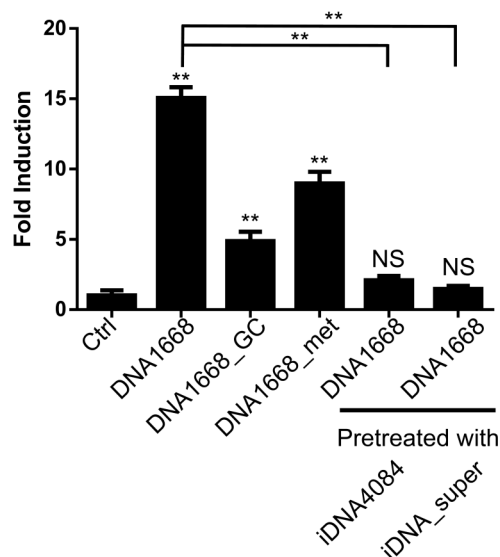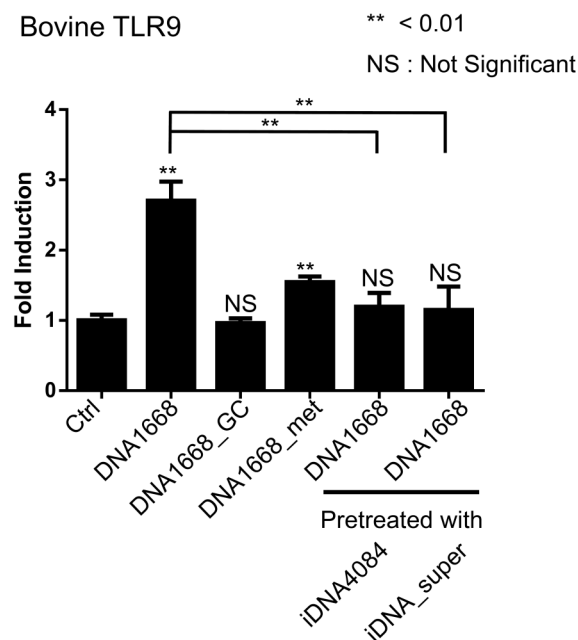
Extended Data Figure 2 | The NF-κB activation experiments.
a, DNA1668_12mer retains agonistic activity to TLR9. The NF-κB activation of wild-type mouse TLR9 induced by DNA1668 (TCCATGACGTTCCTGATGCT), DNA1668_12mer (CATGACGTTCCT), DNA1668 or DNA1668_12mer with a CpG to GpC inversion (DNA1668_GC, DNA1668_12mer_GC). DNAs were all complexed with N-[1-(2,3-Dioleoyloxy)propyl]-N, N, N-trimethylammonium methyl-sulphate (DOTAP) and added at concentration of 1 μM. The activities were analysed with an NF-κB-dependent luciferase reporter assay using HEK293T cells co-expressing mouse TLR9 and mouse Unc93B1. A two-tailed t-test was used to determine the statistical significance of differences between control (Ctrl) and stimulated cells, or between each group. **P < 0.01. Data from three independent experiments are shown. **b**, Horse and bovine TLR9 responses against agonistic and inhibitory DNAs. The NF-κB activation of wild-type horse (left) or bovine (right) TLR9 induced by indicated DNAs. The activities were analysed with an NF-κB-dependent luciferase reporter assay using HEK293T cells co-expressing horse or bovine TLR9 and human Unc93B1. The concentration of agonistic DNAs (DNA1668, DNA1668_GC and DNA1668_met) and inhibitory DNAs (iDNA4084 and iDNA_super) were 10 μM and 1 μM, respectively. Data represent the mean fold induction of NF-κB activity +s.d. (n = 3). A two-tailed t-test was used to determine the statistical significance of differences between control (Ctrl) and stimulated cells, or between each group. **P < 0.01. Data from three independent experiments are shown.

a

EcTLR9_Z-loop unprocessed

b

EcTLR9_Z-loop processed

K_d = 20 µM
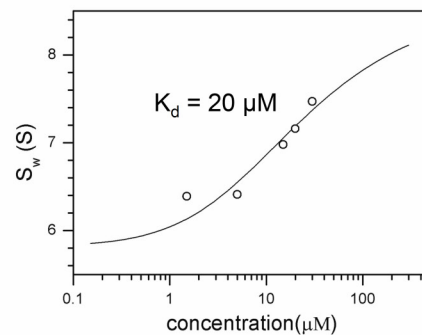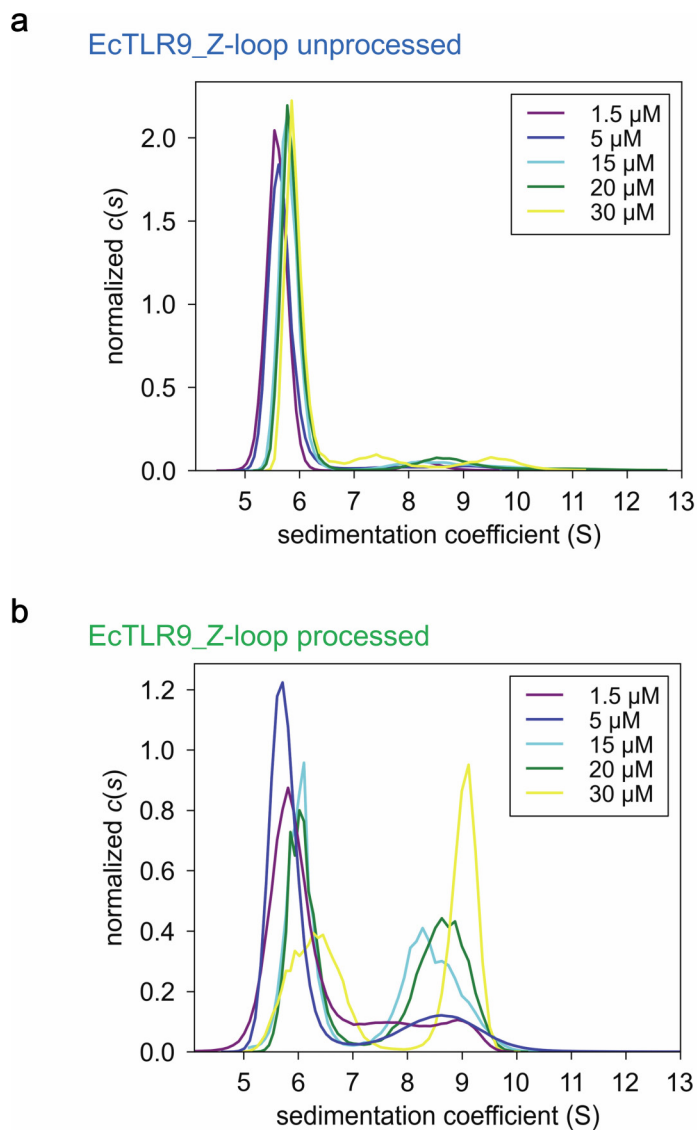
Extended Data Figure 3 | Dimerization interaction of TLR9. a, b, The oligomerization states of EcTLR9 with an unprocessed Z-loop (**a**) or a processed Z-loop (**b**) were analysed by SV–AUC at various concentrations of TLR9–DNA1668_12mer (equimolar). The weight-average sedimentation coefficients (S_w) were plotted against TLR9–DNA1668_12mer concentration to determine the K_d value for the dimerization. The dissociation constant for the dimerization of the processed TLR9 is estimated to be 20 µM.

a EcTLR9/DNA1668_12mer

DNA1668_12 mer
5'-CAT<u>GACGTT</u>CCT-3'

b BtTLR9/DNA1668_12mer

DNA1668_12 mer
5'-CATGACGTTCCT-3'

c EcTLR9/iDNA4084

iDNA4084   5'-<u>CCT</u>GGA<u>TGGG</u>AA-3'

d MmTLR9/iDNA4084

iDNA4084   5'-<u>CCT</u>GGA<u>TGGG</u>AA-3'

e MmTLR9/iDNA_super

iDNA_super   5'-<u>CCTC</u>AATAGGGT<u>GAGGG</u>G-3'
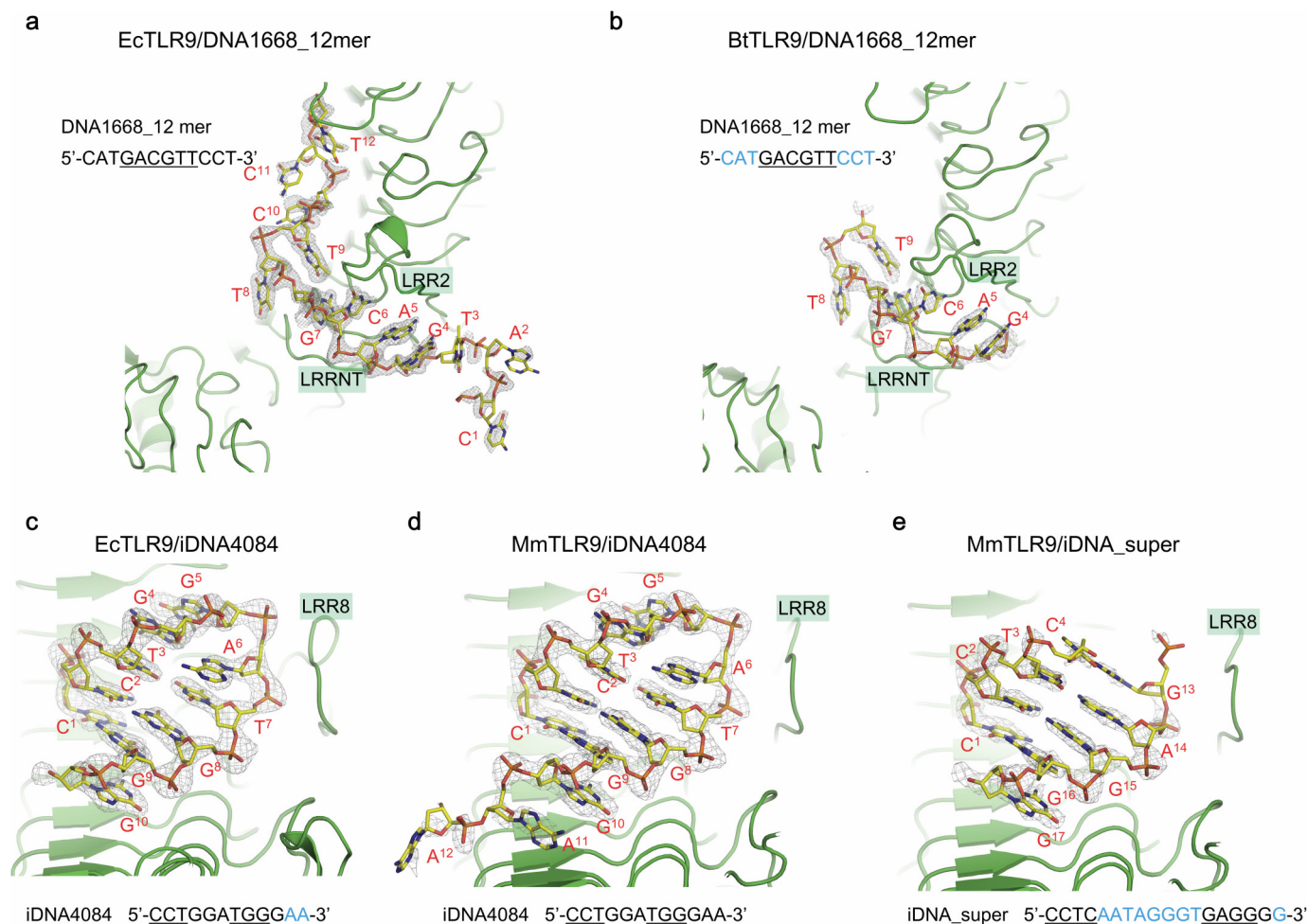
Extended Data Figure 4 | Electron densities of DNA bound to TLR9.
a–e, The F_o–F_c omit difference electron densities of DNA1668_12mer bound to EcTLR9 (**a**) and BtTLR9 (**b**), iDNA4084 bound to EcTLR9 (**c**) and MmTLR9 (**d**), and iDNA_super bound to MmTLR9 (**e**) contoured at the 3.0σ level. The residues coloured blue in the sequence are not visible in the electron density map. The core hexamer could be unambiguously modelled into the continuous electron density map in the EcTLR9–DNA1668_12mer and BtTLR9–DNA1668_12mer complexes, whereas flanking bases were obscure or not visible. The A5–T12 loop connecting the base-paired region of iDNA_super was not visible in the electron density map, whereas the G4–A6 sequence of iDNA4084 was well defined.

a EcTLR9
(EcTLR9/DNA1668_12mer)

b Human TLR8
(Human TLR8/CL097)

c EcTLR9/DNA1668_12mer
Human TLR8/CL097

d EcTLR9/DNA1668_12mer
BtTLR9/DNA1668_12mer

e EcTLR9/DNA1668_12mer
BtTLR9/DNA1668_12mer

f

g EcTLR9/iDN4084
MmTLR9/iDN4084
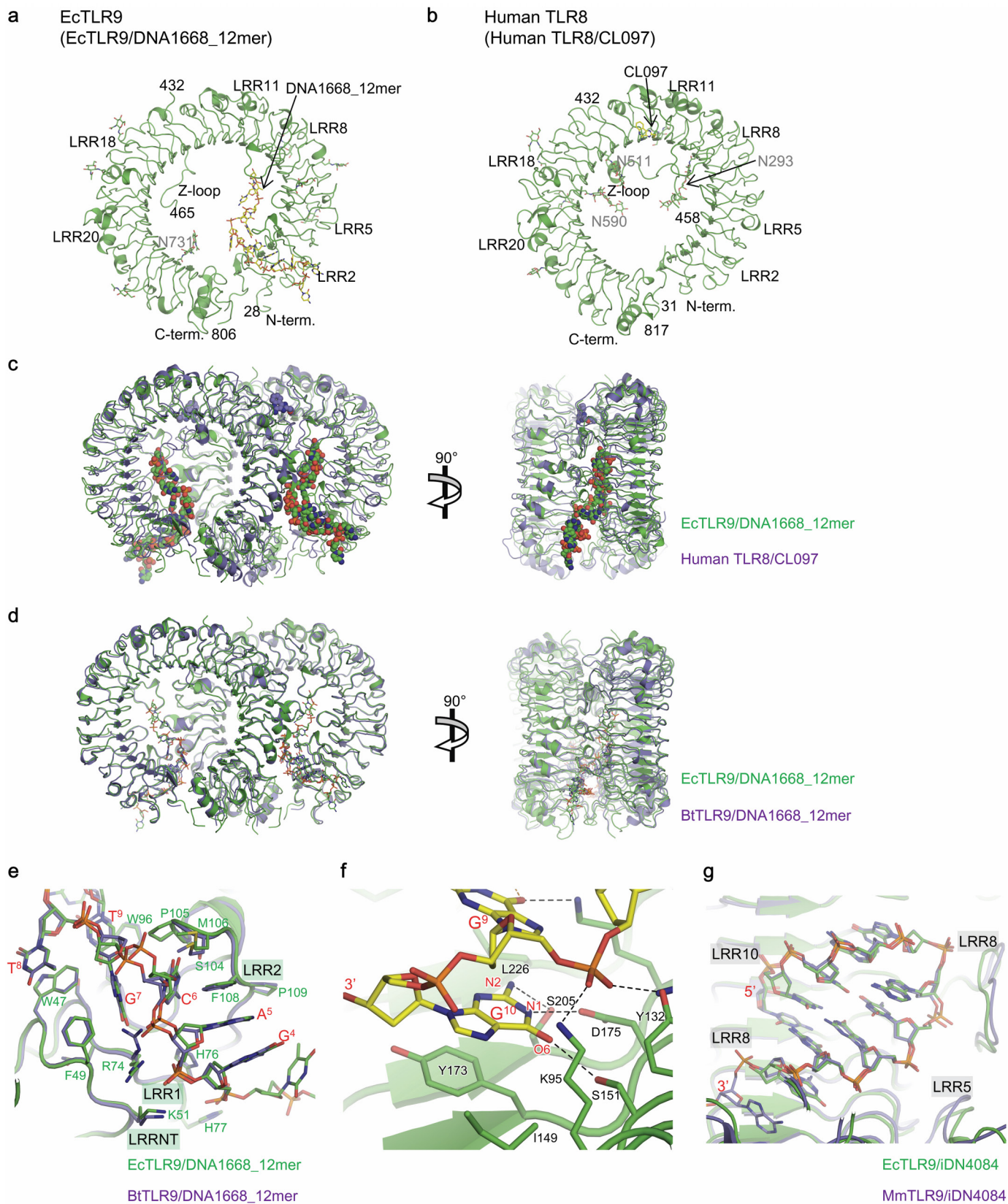
Extended Data Figure 5 | Structures of TLR9. a, Monomer structure of EcTLR9, derived from the EcTLR9–DNA1668_12mer complex. The structure and binding mode of TLR9 are markedly different from those of other CpG-binding proteins[20,31–34]. **b**, Monomer structure of human TLR8, derived from the human TLR8–CL097 (2-(ethoxymethyl)-1H-imidazo[4,5-c]quinolin-4-amine) complex (PDB ID: 3W3J)[16]. The Z-loops in TLR9 and TLR8 are oriented differently with respect to the concave face of TLR and engage in different interactions with it. The latter half of the Z-loop of TLR8 extends towards the N-terminus, whereas the Z-loop of TLR9 extends towards the C-terminus to interact with LRR15–21. TLR8 has three ordered *N*-glycans attached to Asn293, Asn511 and Asn590 that project into the inner space of the ring structure, whereas EcTLR9 has only one *N*-glycan attached to Asn731 that projects inward. As a result, the ring structure of TLR9 has more unoccupied inner space than that of TLR8, an arrangement that is suitable for ligand binding on the concave interior surface. **c**, Superposition of the overall ligand-induced dimer structures of EcTLR9 (DNA1668_12mer complex, green) and human TLR8 (CL097 complex, purple) by PyMol[28], yielding an root-mean-square deviation (r.m.s.d.) value of 2.3 Å. **d**, Superposition of the overall dimer structures of EcTLR9 (DNA1668_12mer complex, green) and

BtTLR9 (DNA1668_12mer complex, purple) by PyMol[28], yielding an r.m.s.d. value of 0.7 Å. **e**, Magnified view of the CpG-binding groove of EcTLR9 and BtTLR9. The amino acid at position 109 was proline for human and serine for mouse (rodents). From the structure, Pro109 made a van der Waals contact with A at the –1 position, but its contact was somewhat close. If A is changed to T, this contact is weakened. Serine at position 109 would accommodate a larger base. **f**, Magnified view of the G10 of iDNA4084 recognition by TLR9 (EcTLR9–iDNA4084 complex). G10 of iDNA4084 makes three hydrogen bonds with TLR9: N2, N1 and O6 atoms of G10 with the side chains of Ser205, Asp175 and Ser151, respectively. To examine the functional importance of this base, we substituted it with other bases and examined the binding affinity by ITC. The affinity of TLR9 for G10A (K_d = 6 nM) was reduced from that of the original sequence (K_d = 3 nM), but TLR9 exhibited much lower affinity for DNAs with a pyrimidine at this position (K_d = 41 nM for G10C and K_d = 76 nM for G10T), suggesting that this position favours purine over pyrimidine. **g**, Superposition of iDNA4084 bound to EcTLR9 (green) and MmTLR9 (purple). The binding mode of iDNA4048 is perfectly conserved between EcTLR9 and MmTLR9.
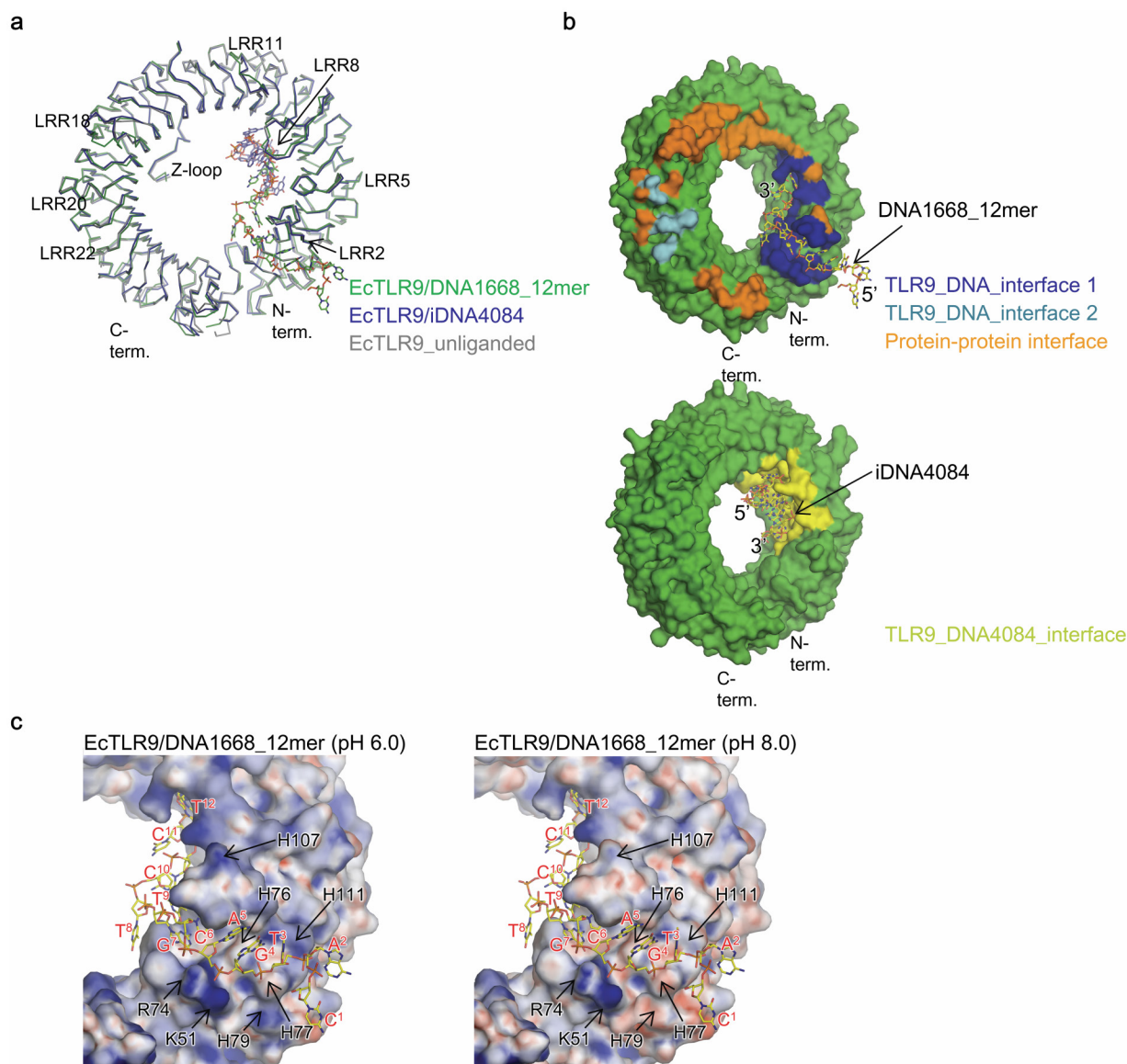
Extended Data Figure 6 | Binding interfaces for agonistic and inhibitory DNA. **a**, Superposition of the structures of unliganded (grey), DNA1668_12mer-bound (green), and iDNA4084-bound (blue), EcTLR9. TLR9 and DNA are shown in Cα-trace and stick representation, respectively. No significant conformational change was observed upon agonistic DNA binding, as indicated by the small r.m.s.d. value of 0.8 Å between EcTLR9 (unliganded) and the EcTLR9–DNA1668_12mer complex. Instead, the conversion of EcTLR9 into the activated form appears to involve local conformational changes in the loop regions of LRR8, LRR11 and LRR18, all of which are involved in formation of the dimer. No significant structural change was induced by binding to iDNA4084, as indicated by the small r.m.s.d. values of 0.49 Å and 0.45 Å between the unliganded and iDNA4084-bound forms of EcTLR9 and MmTLR9, respectively. **b**, Surface representations of EcTLR9 structures in the DNA1668_12mer (upper) and iDNA4084 complexes (lower). The protein–protein interface, TLR9_DNA1668_12mer interfaces 1 and 2, and TLR9_iDNA4084 interface are shown in orange, blue, cyan and yellow,

respectively. The bound DNAs are shown in stick representation with their 5′ and 3′ ends indicated. DNA1668_12mer buries approximately 1,136 Å² and 294 Å² of the accessible surface area of TLR9 and TLR9*, respectively, suggesting that the N-terminal binding site of TLR9 for DNA1668_12mer makes a relatively larger contribution to binding. The binding site for iDNA partially overlaps with the binding site for agonistic DNA. Specifically, LRR4 and LRR5 are both involved in the binding sites for agonistic DNA and iDNAs, although the binding modes of DNA1668_12mer and iDNA4084 are completely different: Arg152 (LRR4), Tyr179 and Lys181 (LRR5) interact with the phosphate of C11 of DNA1668_12mer but also with G8 of iDNA4084. **c**, Electrostatic potential map of DNA-binding region. The map was calculated at basic and acidic conditions by PyMol[28]. Surface colours represent the potential from $-20\,k_\mathrm{B}T/e$ (red) to $20\,k_\mathrm{B}T/e$ (blue), where k_B is the Boltzman constant and T is the absolute temperature. The DNA molecule is shown as a stick model.
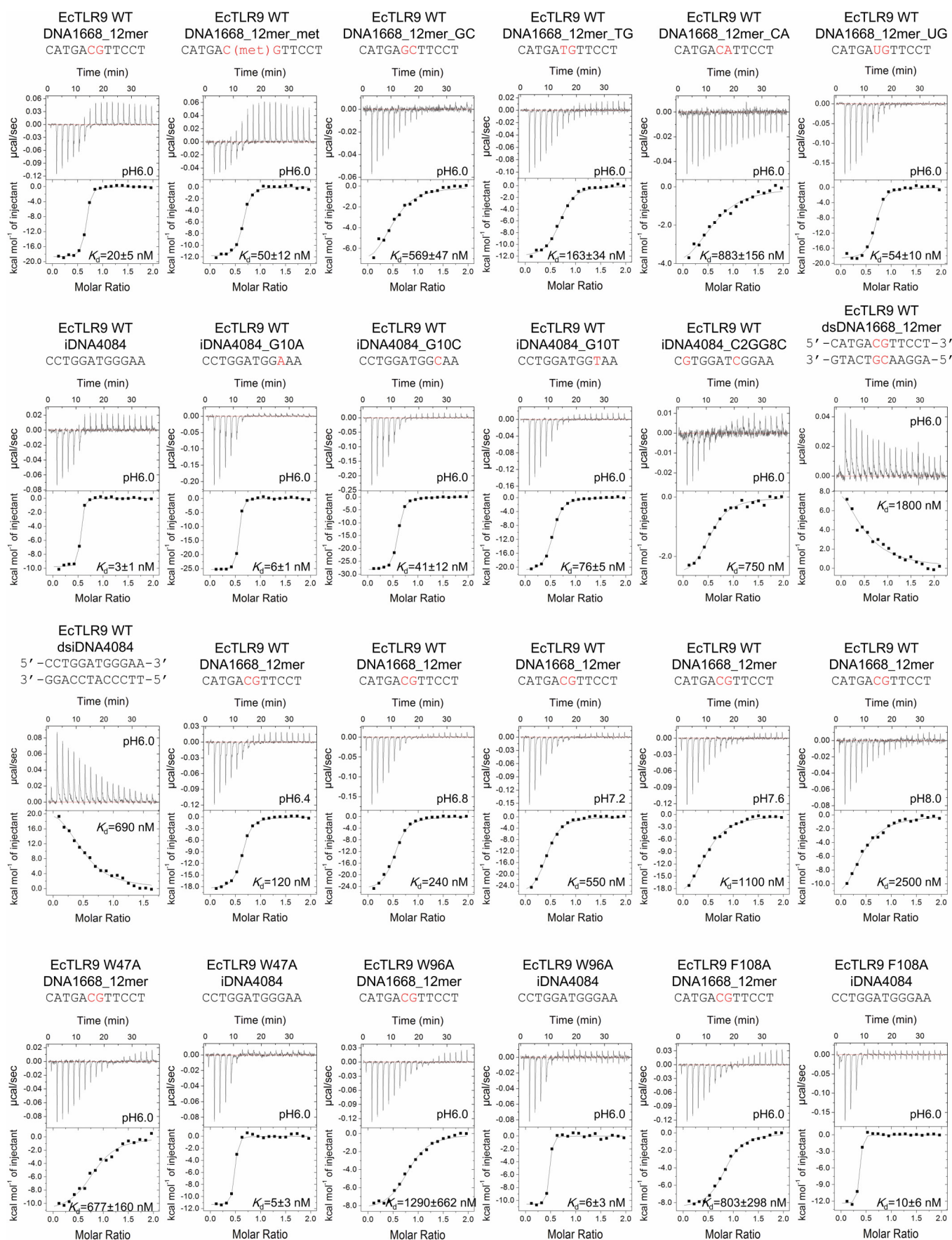
Extended Data Figure 7 | ITC thermograms for ITC data (related to Extended Data Table 2). Representative ITC thermograms for the ITC data are shown with their pH condition and K_d values for EcTLR9.
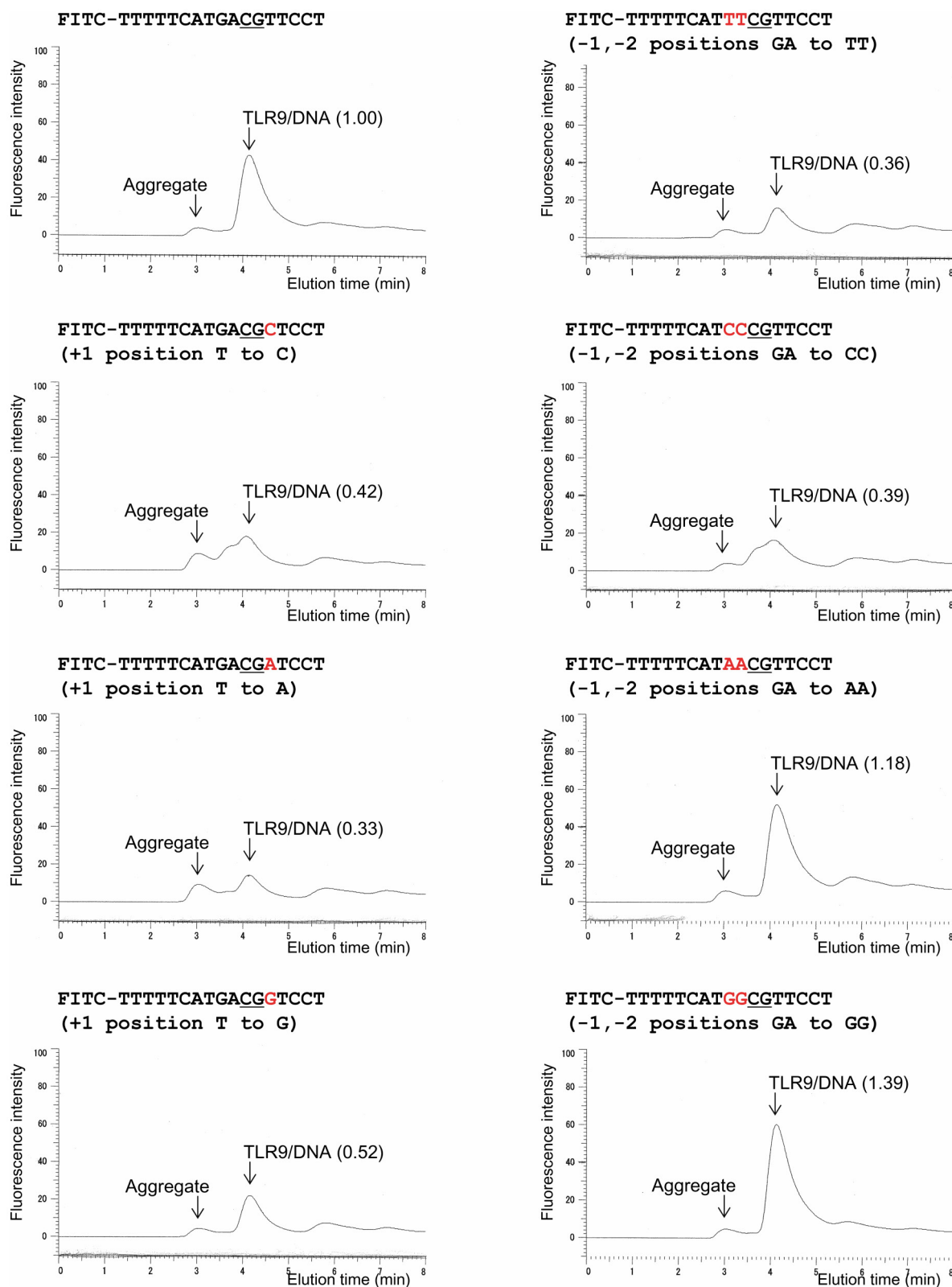
Extended Data Figure 8 | Significance of the consensus sequence in the CpG-DNA. Gel-filtration chromatography of EcTLR9 with FITC-labelled DNA. DNA binding to TLR9 was monitored by FITC-fluorescence (excitation 495 nm, emission 520 nm). The parenthesized values indicate the ratios of the fluorescence peak height of the derivative DNA to the original DNA containing the consensus sequence of GACGTT (top left). In each experiment, 0.5 μM EcTLR9 (total volume of 45 μl) with DNA (equimolar) was injected into a Superdex 200 Increase 5/150 GL (GE healthcare) gel-filtration column. The running buffer was 10 mM 2-morpholinoethanesulfonic acid and 250 mM NaCl at pH 5.5. DNAs used in the analyses are shown in each panel. Bases that are changed from the original sequence are highlighted in red. Conversion of the purine–purine sequence (GA) at the –1 and –2 positions of the CpG motif (underlined) to AA and GG resulted in DNAs with affinities similar to the wild-type DNA, but conversion to a pyrimidine–pyrimidine sequence (TT and CC) weakened the affinity, demonstrating that a purine–purine sequence is preferable at these positions. Conversion of the base at the +1 position into C, A or G led to weaker binding, suggesting that T is preferable at the +1 position, although T is not specifically recognized.

Extended Data Table 1 | Crystallization, data collection and refinement statistics

	EcTLR9 (Unliganded)	EcTLR9 (DNA1668_12mer)	EcTLR9 (iDNA4084)	EcTLR9 (DNA1668_12mer)	BtTLR9 (Unliganded)	MmTLR9 (iDNA4084) Form1	MmTLR9 (iDNA4084) Form2	MmTLR9 (iDNA_super)
Crystallization								
Protein solution	4.0 mg/ml EcTLR9 10 mM Tris pH 7.5 0.15 M NaCl	4.3 mg/ml EcTLR9 10 mM Tris pH 7.5 0.35 M NaCl 140 μM DNA1668_12mer	5.8 mg/ml EcTLR9 10 mM Tris pH 7.5 0.35 M NaCl 130 μM iDNA4084	6.8 mg/ml BtTLR9 10 mM Tris pH 7.5 0.35 M NaCl 150 μM DNA1668_12mer	5.1 mg/ml MmTLR9 10 mM Tris pH 7.5 0.15 M NaCl	5.1 mg/ml MmTLR9 10 mM Tris pH 7.5 0.15 M NaCl 110 μM iDNA4084	5.1 mg/ml MmTLR9 10 mM Tris pH 7.5 0.15 M NaCl 110 μM iDNA4084	5.6 mg/ml MmTLR9 10 mM Tris pH 7.5 0.35 M NaCl 120 μM iDNA_super
Reservoir solution	10–14% (w/v) PEG3350 0.2 M AS 0.1 M Na acetate pH 4.5 10% (v/v) ethylene glycol	1% (w/v) PEG 4000 1% (v/v) 2-propanol 0.2 M Na acetate 0.1 M Na citrate pH5.8 5% (v/v) DMSO	10–14% (w/v) PEG 3350 0.2 M AS 0.1 M Na acetate pH 4.5 10% (v/v) ethylene glycol	3–5% (w/v) PEG 4000 0.2 M Na acetate 0.1 M Na citrate pH 5.5 15% (v/v) DMSO	6–8% (w/v) PEG 8000 0.4 M AS 0.1 M Tris pH 8.0 10% (v/v) DMSO	6–8% (w/v) PEG 8000 0.4 M AS 0.1 M Tris pH 8.0 10% (v/v) DMSO	6–8% (w/v) PEG 8000 0.4 M AS 0.1 M Tris pH 8.0 10% (v/v) DMSO	6–8% (w/v) PEG 8000 0.4 M AS 0.1 M Tris pH 8.0 10% (v/v) DMSO
Cryoprotectant solution	15% (w/v) PEG3350 0.2 M AS 0.1 M Na acetate pH 4.5 20% (v/v) ethylene glycol	5% (w/v) PEG 4000 5% (v/v) 2-propanol 0.15 M Na acetate 0.25 M NaCl 0.1 M Na citrate pH5.8 10% (v/v) DMSO 20% (v/v) glycerol 1/10 vol. of Silver Bullet A6	15% (w/v) PEG 3350 0.2 M AS 0.1 M Na acetate pH 4.5 20% (v/v) ethylene glycol	8% (w/v) PEG 4000 0.2 M Na acetate 0.1 M Na citrate pH 5.5 15% (v/v) DMSO 25% (v/v) glycerol	10% (w/v) PEG 8000 0.4 M AS 0.1 M NaCl 0.1 M Tris pH 8.0 10% (v/v) DMSO 15% (v/v) glycerol	10% (w/v) PEG 8000 0.4 M AS 0.1 M NaCl 0.1 M Tris pH 8.0 10% (v/v) DMSO 15% (v/v) glycerol	10% (w/v) PEG 8000 0.4 M AS 0.1 M NaCl 0.1 M Tris pH 8.0 10% (v/v) DMSO 15% (v/v) glycerol	10% (w/v) PEG 8000 0.4 M AS 0.1 M NaCl 0.1 M Tris pH 8.0 10% (v/v) DMSO 15% (v/v) glycerol
Data collection								
Space group	$C2$	$P1$	$C2$	$C2$	$P2_12_12_1$	$P2_12_12_1$	$C2$	$C2$
Cell dimensions								
a, b, c (Å)	98.1, 118.8, 77.8	68.8, 81.6, 81.5	117.6, 117.1, 69.6	115.8, 113.1, 67.4	71.3, 119.3, 130.4	71.3, 123.9, 130.4	136.4, 130.1, 71.3	137.3, 129.7, 71.4
α, β, γ (°)	90, 108.7, 90	87.5, 83.5, 86.0	105.5	100.6			91.5	92.2
Resolution (Å)	36.8–2.40 (2.53–2.40)*	50.0–1.60 (1.66–1.60)	67.1–2.75 (2.90–2.75)	48.1–2.38 (2.51–2.38)	48.11–1.96 (2.06–1.96)	55.9–2.25 (2.37–2.25)	57.5–2.33 (2.45–2.33)	57.6–2.25 (2.37–2.25)
R_{sym}	0.162 (0.611)	0.078 (0.537)	0.143 (0.530)	0.113 (0.598)	0.113 (0.486)	0.120 (0.836)	0.084 (0.453)	0.072 (0.449)
$I/\sigma I$	6.1 (2.1)	9.0 (1.8)	7.2 (2.3)	7.7 (2.1)	7.9 (2.8)	9.6 (1.9)	8.5 (2.1)	10.2 (2.2)
Completeness (%)	100.0 (100.0)	95.7 (84.4)	99.9 (99.9)	99.8 (99.3)	98.6 (98.4)	99.9 (99.8)	99.7 (99.3)	97.2 (94.0)
Redundancy	3.6 (3.6)	3.8 (2.8)	3.5 (3.2)	3.6 (3.5)	4.3 (3.8)	6.9 (5.8)	3.5 (3.1)	3.7 (3.3)
Refinement								
Resolution (Å)	36.9–2.4	47.5–1.60	58.6–2.75	44.6–2.38	44.6–1.96	35.6–2.25	38.9–2.33	36.3–2.25
No. reflections	31,314	210,033	22,352	32,294	79,109	55,735	53,154	57,594
R_{work} / R_{free}	19.7 / 25.0	18.4 / 22.0	20.2 / 25.9	21.7 / 27.4	22.4 / 26.4	22.9 / 27.7	20.8 / 24.8	22.1 / 25.4
No. atoms								
Protein	5,909	12,022	5,909	5,628	5,858	5,757	5,728	5,750
DNA		478	206	124		227	248	183
Water	112	1,595	10	31	236	77	150	102
B-factors								
Protein	30.2	20.3	43.8	50.5	37.2	57.4	49.1	51.7
DNA		39.2	38.4	57.5		46.6	37.6	66.0
Water	29.0	32.6	24.7	41.9	37.7	52.7	42.2	46.8
R.m.s deviations								
Bond lengths (Å)	0.006	0.014	0.009	0.006	0.006	0.005	0.005	0.004
Bond angles (°)	1.25	1.68	1.44	1.17	1.17	0.93	1.00	0.99

Each data set was collected with one crystal.

* Highest resolution shell is shown in parenthesis.

† Silver Bullet is a crystallization screen kit for biological macromolecules (Hampton Research Inc.). Silver Bullet A6 contains 0.33% (w/v) 2,6-naphthalenedisulfonic acid disodium salt, 0.33% (w/v) 2,7-naphthalenedisulfonic acid disodium salt, 0.33% (w/v) anthraquinone-2,6-disulfonic acid disodium salt and 0.02 M Na-HEPES pH 6.8.

Extended Data Table 2 | ITC results

Protein	pH	DNA	Sequence 5' →3'	K_d (nM)	ΔH (kcal/mol)	ΔS (cal/mol/deg)	N	No. of experiment
Horse TLR9_WT	6.0	DNA1668_12mer	CATGACGTTCCT	20 ± 5	-18.8 ± 1.5	-26.9 ± 5.9	0.5	4
	6.0	DNA1668_12mer_met	CATGAC(met)GTTCCT	50 ± 12	-11.9 ± 0.6	-6.5 ± 2.1	0.6	5
	6.0	DNA1668_12mer_GC	CATGAGCTTCCT	570 ± 47	-7.7 ± 0.8	2.8 ± 2.7	0.6	3
	6.0	DNA1668_12mer_TG	CATGATGTTCCT	160 ± 34	-12.4 ± 0.5	-10.5 ± 2.2	0.7	3
	6.0	DNA1668_12mer_CA	CATGACATTCCT	880 ± 160	-4.1 ± 0.6	14.1 ± 2.3	0.7	3
	6.0	DNA1668_12mer_UG	CATGAUGTTCCT	54 ± 10	-17.4 ± 2.1	-24.9 ± 7.4	0.7	4
	6.0	iDNA4084	CCTGGATGGGAA	3 ± 1	-9.6 ± 0.2	6.7 ± 1.1	0.5	3
	6.0	iDNA4084_G10A	CCTGGATGGAAA	6 ± 1	-25.1 ± 0.2	-46.4 ± 1.1	0.5	3
	6.0	iDNA4084_G10C	CCTGGATGGCAA	41 ± 12	-27.4 ± 1.2	-57.9 ± 4.1	0.4	3
	6.0	iDNA4084_G10T	CCTGGATGGTAA	76 ± 5	-21.7 ± 0.3	-40.1 ± 1.1	0.4	3
	6.0	iDNA4084_C2GG8C	CGTGGATCGGAA	750	-2.8	18.7	0.5	1
	6.0	dsDNA1668_12mer	5'-CATGACGTTCCT-3' 3'-GTACTGCAAGGA-5'	1800	13.4	71.3	0.5	1
	6.0	dsiDNA4084	5'-CCTGGATGGGAA-3' 3'-GGACCTACCCTT-5'	690	24.7	111.0	0.5	1
Horse TLR9_WT	6.4	DNA1668_12mer	CATGACGTTCCT	120	-22.0	41.9	0.6	1
	6.8	DNA1668_12mer	CATGACGTTCCT	240	-29.1	-67.3	0.5	1
	7.2	DNA1668_12mer	CATGACGTTCCT	550	-36.5	-93.9	0.4	1
	7.6	DNA1668_12mer	CATGACGTTCCT	1100	-30.2	-74.0	0.5	1
	8.0	DNA1668_12mer	CATGACGTTCCT	2500	-26.0	-61.5	0.5‡	1
Horse TLR9_W47A	6.0	DNA1668_12mer	CATGACGTTCCT	680 ± 160	-11.6 ± 2.2	-10.5 ± 7.5	1.2	3
	6.0	iDNA4084	CCTGGATGGGAA	5 ± 3	-11.4 ± 0.2	-0.1 ± 1.9	0.4	3
Horse TLR9_W96A	6.0	DNA1668_12mer	CATGACGTTCCT	1300 ± 660	-11.4 ± 2.9	-10.9 ± 11.2	0.9	3
	6.0	iDNA4084	CCTGGATGGGAA	6 ± 3	-12.7 ± 2.4	-4.8 ± 6.8	0.4	3
Horse TLR9_F108A	6.0	DNA1668_12mer	CATGACGTTCCT	800 ± 300	-8.1 ± 3.4	0.7 ± 12.2	1.0	3
	6.0	iDNA4084	CCTGGATGGGAA	10 ± 6	-10.5 ± 2.0	1.7 ± 6.8	0.4	3

LETTER

The octahaem MccA is a haem *c*–copper sulfite reductase

Bianca Hermann[1], Melanie Kern[2], Luigi La Pietra[2], Jörg Simon[2] & Oliver Einsle[1,3]

The six-electron reduction of sulfite to sulfide is the pivot point of the biogeochemical cycle of the element sulfur[1]. The octahaem cytochrome *c* MccA (also known as SirA) catalyses this reaction for dissimilatory sulfite utilization by various bacteria. It is distinct from known sulfite reductases because it has a substantially higher catalytic activity and a relatively low reactivity towards nitrite. The mechanistic reasons for the increased efficiency of MccA remain to be elucidated. Here we show that anoxically purified MccA exhibited a 2- to 5.5-fold higher specific sulfite reductase activity than the enzyme isolated under oxic conditions[2]. We determined the three-dimensional structure of MccA to 2.2 Å resolution by single-wavelength anomalous dispersion. We find a homotrimer with an unprecedented fold and haem arrangement, as well as a haem bound to a $CX_{15}CH$ motif[3]. The heterobimetallic active-site haem 2 has a Cu(I) ion juxtaposed to a haem *c* at a Fe–Cu distance of 4.4 Å. While the combination of metals is reminiscent of respiratory haem–copper oxidases, the oxidation-labile Cu(I) centre of MccA did not seem to undergo a redox transition during catalysis. Intact MccA tightly bound SO_2 at haem 2, a dehydration product of the substrate sulfite that was partially turned over due to photoreduction by X-ray irradiation, yielding the reaction intermediate SO. Our data show the biometal copper in a new context and function and provide a chemical rationale for the comparatively high catalytic activity of MccA.

Sulfate-reducing microbes couple the reduction of sulfate, SO_4^{2-}, and sulfite, SO_3^{2-}, to the generation of a proton motive force to sustain organismic growth. A single enzymatic step is required for the six-electron

reduction of sulfite—present at physiological pH ($pK_a = 6.97$) in equilibrium with the monovalent hydrogen sulfite anion, HSO_3^-—to sulfide, either using the tetrapyrrol cofactor sirohaem in assimilatory or dissimilatory sulfite reductases[1], or a haem *c* as a side reactivity of the pentahaem cytochrome *c* nitrite reductase NrfA[4,5]. Reported specific sulfite reduction rates for these enzymes are low, with turnover number (k_{cat}) values ranging from 0.2 to 2 s^{-1} (refs 4, 6, 7). All known sulfite reductases also catalyse the six-electron reduction of nitrite, NO_2^-, to ammonium, NH_4^+, and in the case of NrfA, a single-point variant at the active site loses nitrite reductase activity, while leaving its basal sulfite reductase activity unchanged[4]. The genome of the Epsilonproteobacterium *Wolinella succinogenes* does not encode a sirohaem sulfite reductase and the *nrfA* gene is not induced during sulfite respiration[2]. Instead, sulfite is reduced by the octahaem *c*-type cytochrome MccA, with sulfide as the sole product[2,8]. The reported activity of MccA exceeded the activity levels reported for other sulfite-reducing enzymes by far, but the mechanistic basis for this higher efficiency remained unclear[1,2]. MccA belongs to the genetically diverse family of multihaem *c* enzymes and has eight haem groups covalently attached to conserved haem-binding motifs in the peptide sequence[2,3,9]. Multihaem cytochrome *c* enzymes show a high conservation of haem group arrangements, but not of sequence, with recurring haem-packing motifs that result in either a parallel or a perpendicular packing of two of the moieties[10,11]. They catalyse complex multi-electron redox reactions and bind their substrates through the free electron pairs of a heteroatom to a free coordination
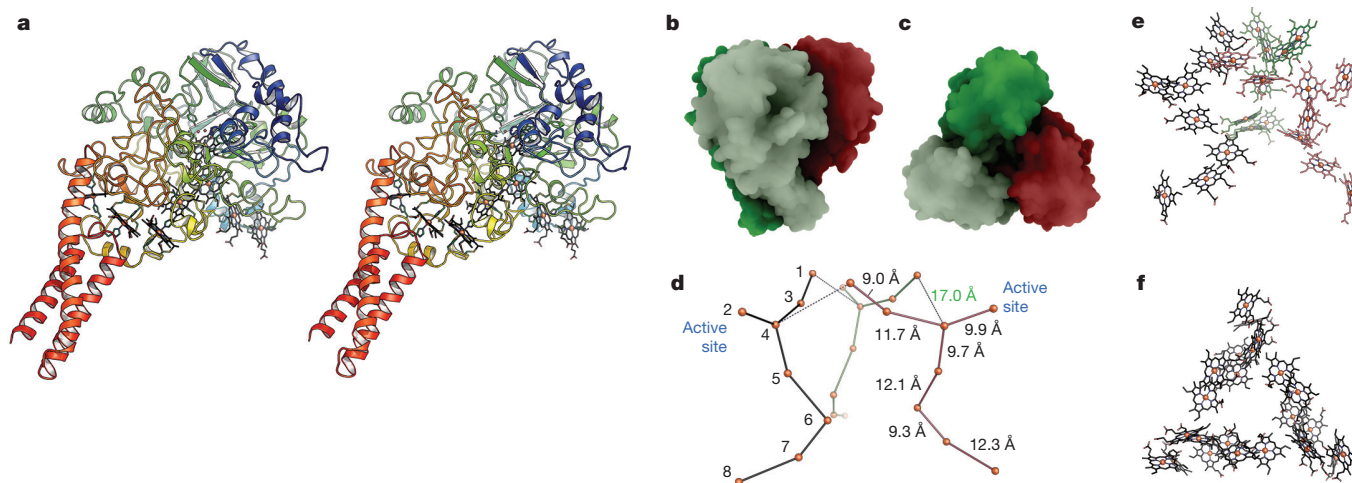

Figure 1 | Structure of *W. succinogenes* MccA. a, Stereo representation of the MccA monomer coloured from blue (N terminus) to red (C terminus), highlighting the prominent trimerization helices. **b**, Surface representation of the functional trimer of MccA. The trimerization helices are at the bottom and the substrate access funnel in the globular domain is well visible. **c**, Top view of the trimer, coloured as in **b**. **d**, Arrangement of the 24 iron atoms within the MccA trimer. The haem numbering in the left monomer follows the

sequence of attachment to the protein chain. The active sites are located at haem 2 of each subunit. Fe–Fe distances are given for one monomer, and the closest Fe–Fe distance between monomers is between haems 1 and 4′ at 17 Å. **e**, Cofactor arrangement. The 24 haem groups form a complex pattern of interacting parallel and perpendicular packing motifs. **f**, A top view of the haem groups emphasizes the trimeric cofactor arrangement.

[1]Lehrstuhl Biochemie, Institut für Biochemie, Albert-Ludwigs-Universität Freiburg, Albertstrasse 21, 79104 Freiburg, Germany. [2]Microbial Energy Conversion & Biotechnology, Department of Biology, Technische Universität Darmstadt, Schnittspahnstrasse 10, 64287 Darmstadt, Germany. [3]BIOSS Centre for Biological Signalling Studies, Schänzlestrasse 1, 79104 Freiburg, Germany.
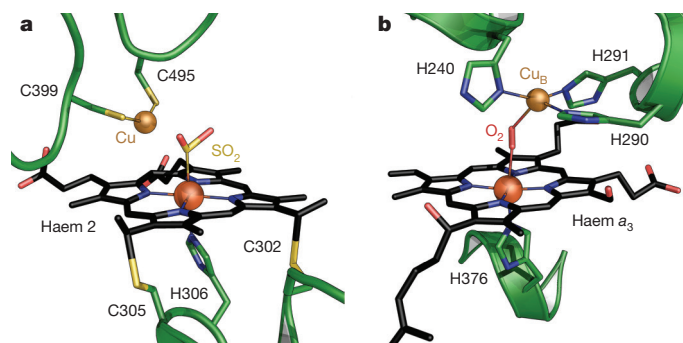
Figure 2 | Haem–copper active sites in respiratory enzymes.
a, *W. succinogenes* MccA binds its substrate sulfite in the dehydrated form, SO_2, at the distal axial position of haem 2. At 3.2 Å distance from the sulphur atom, a Cu(I) ion is nearly linearly coordinated by residues C399 and C495. **b**, In respiratory haem–copper oxidases, Cu_B is a redox-active species liganded by three histidine residues and juxtaposed to a haem a_3 moiety. The arrangement, with a Fe–Cu distance of 4.9 Å, is optimized to bind O_2 and peroxide in a bridging fashion (Protein Data Bank accession 3ABM).

position at an active-site haem group. Electrons are then provided or accepted by the tightly coupled chain of haem groups[12]. To understand the functional specifications that enable the increased sulfite reductase activity of MccA, we isolated the enzyme from *W. succinogenes* and determined its three-dimensional structure to 2.2 Å resolution (Extended Data Table 1). The active site of MccA contained an additional metal ion that was depleted during oxic isolation procedures. This ion was identified as Cu(I), making MccA a new type of heterobimetallic oxidoreductase, reminiscent of the membrane-integral respiratory haem–copper oxidases.

Physiological sulfite respiration occurs under anaerobic conditions, but Strep-tagged MccA from *W. succinogenes* was previously isolated only in the presence of dioxygen[2,3]. Here, we produced MccA both under strict exclusion of dioxygen (form I) and under oxic conditions (form II) for comparison. First, crystals were obtained for form II at <5 p.p.m. O_2, and whereas the anoxically prepared form I crystallized as it was isolated, the crystallization of MccA II required reduction with dithionite or the addition of sulfite. The protein was modelled from amino acids M31 to S689, with disorder only in the termini (Fig. 1a). Aside from the relevant changes at the active site (discussed later), no major differences were observed between the individual structures. The biological assembly of MccA is a trimer with approximate dimensions of $90 \times 90 \times 90$ Å³ (Fig. 1b, c), and the monomer can be subdivided into two domains, with a globular amino-terminal domain that contains few secondary structure elements. This domain holds eight covalently attached haem groups, one of which (haem 2) forms the active site (Fig. 1d, e). Haem 8 is attached to a non-canonical binding motif ($CX_{15}CH$) within the peptide chain, as previously shown by mass spectrometry[3]. The carboxy-terminal domain consists of three long α-helices that form an extended multimerization interface, similar to that in the pentahaem NrfA[10,12,13] and the octahaem proteins hydroxylamine oxidoreductase (HAO) from *Nitrosomonas europaea*[14] and nitrite reductase from *Thioalkalivibrio nitratireducens*[15]. In all these cases, the positioning of these helices and the resulting multimers are different, while the relative orientation of the haem groups is conserved (Fig. 1a, d and Extended Data Fig. 1). The haems of MccA line up at the inner periphery of the trimer, allowing access only to the active-site haem 2 situated almost in the centre of the domain, and to haem 8, the likely electron entry site (Fig. 1d, f). Among cytochromes of known structure, the haem groups align best with those of the monomeric octahaem tetrathionate reductase (OTR) from *Shewanella oneidensis*[16], in which haem 2 forms the active site in both cases (Extended Data Fig. 1e). Fe–Fe distances range between 9.0 and 9.9 Å for the parallel packing motifs and between 11.7 and 12.4 Å for the perpendicular motifs, allowing for high electron transfer rates. Between two monomers of the trimer, haem 1 and haem 4′ are
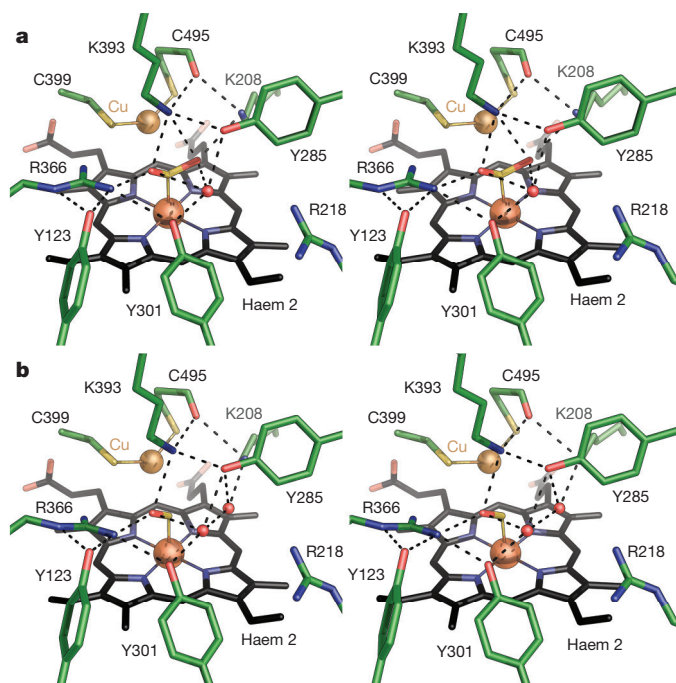

Figure 3 | Sulfite reduction at haem 2 of MccA. a, Sulfite binds to the resting state of the enzyme, but the active-site architecture shifts the reversible dehydration equilibrium to $SO_2 + H_2O$. **b**, Reduction by two electrons occurs through photoreduction in the X-ray beam and leads to release of a second H_2O after protonation, with SO remaining bound to the active site. A tight network of hydrogen bonds surrounds the bimetal centre and the bound substrate, holding the reaction products in place.

at the closest Fe–Fe distance of 17.0 Å, with 9.3 Å separating the porphyrin edges (Fig. 1d). All haems are *bis*-histidinyl coordinated, with the exception of the penta-coordinate active-site haem 2 that has the distal axial position available for ligand binding. In the $CX_{15}CH$ motif of haem 8, the extended region between the two cysteine residues forms a loop with a short helical turn, in direct vicinity to another loop harbouring the only non-proline *cis* peptide in the enzyme, between residues G508 and F509 (Extended Data Fig. 2). Its formation might require the essential peptidyl isomerase MccB[2], and it is presumed to be a prerequisite for correct folding of the loop in the maturation process of haem 8, which is likely to be attached by the dedicated cytochrome *c* synthase CcsA1 (refs 17–19). The structure of the $CX_{15}CH$ haem *c* binding motif disrupts the general parallel/perpendicular haem stacking sequence, and rotates the haem out of plane, possibly to optimize the interaction with the putative electron donor, the iron–sulfur protein MccC (Extended Data Fig. 1a)[2].

The active-site haem 2 is bound to a canonical CXXCH motif with H306 as a proximal axial ligand (Fig. 2a). An unprecedented feature of MccA is the presence of two cysteine residues, C399 and C495, juxtaposed at the distal side of the active-site cavity. In the initial structure of form II of the enzyme, an additional electron density maximum was observed between the cysteine thiolates, coordinated in an extended geometry at distances of 2.1 Å and 2.2 Å, respectively, with an S–X–S angle between 140° and 145°. This mode of coordination was suggestive of Cu(I), so that in a fully occupied form I crystal we used anomalous double-difference ($\Delta\Delta_{ano}$) electron density maps collected at the X-ray absorption edges of several third-row elements (Co, Ni, Cu, Zn, As) to confirm unambiguously the identity of the atom as copper (Extended Data Fig. 3). In form II MccA this site was only partially occupied, resulting in a weaker maximum for the electron density peak and a mixture of two conformations with partial formation of a disulfide bond between C399 and C495 (Extended Data Fig. 4). The oxidative decomposition of the haem–copper site is not reversible by the reduction or
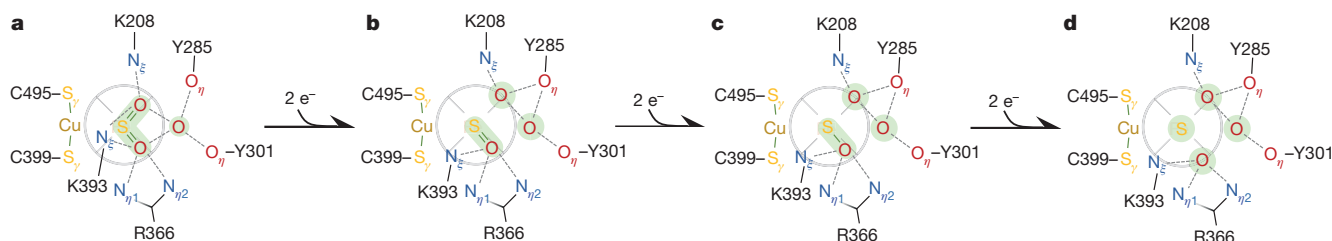
Figure 4 | Mechanism of stepwise sulfite reduction by MccA. a–d, The schemes represent a view onto the plane of haem 2. **a,** Sulfite is bound at an oxidation state of S^{+IV} and immediately dehydrated. All oxygen atoms are fixed in a tight hydrogen-bonding network (Fig. 3a). **b,** Upon transfer of two electrons to yield an S^{+II} state, a second oxygen atom is released as water, but held within the active-site cavity (Fig. 3b). **c,** A further two-electron reduction leads to the S^0 state, weakening the remaining S–O bond. **d,** With the final transfer of two electrons, the state of sulfide (S^{-II}) is reached and a third water is released. In all intermediates, protons are omitted for clarity.

addition of Cu(I) salts, indicating that its assembly requires further factors. One of these may be an orthologue of the copper chaperone NosL, which is encoded by a gene that forms part of the *mcc* gene cluster[1,2,8]. The geometry of the copper site suggests a preference for Cu(I), and metal determinations by inductively coupled plasma emission mass spectrometry (ICP-MS) indeed confirmed that forms I and II of MccA differ markedly in copper content. The high-copper form I, with an average Cu content of 0.7–1.0 per monomer, also showed higher catalytic activity ($k_{cat} = 200.6 \, \text{s}^{-1}$) than form II, with an average of 0.2–0.5 Cu per monomer ($k_{cat} = 99.5 \, \text{s}^{-1}$; Extended Data Fig. 5), in good agreement with the Cu occupancy observed in the crystal structures. The active centre of MccA thus consists of an unprecedented haem *c*–Cu(I) arrangement, with a metal–metal distance of 4.4 Å (Fig. 2a). This is a slightly shorter distance than the 4.9 Å observed in the haem a_3–Cu_B site of cytochrome *c* oxidase (Fig. 2b), the only other haem–copper enzyme known, where the diatomic peroxide intermediate binds in a linear μ–η:η bridging mode[20].

MccA reduced sulfite, but not arsenate, selenate, selenite, hydroxylamine, hydrazine, fumarate, nitrate, thiosulfate, tetrathionate, polysulfide or Fe(III)[2,3]. Nitrite was reduced only very slowly to ammonium, with a k_{cat} of 16.1 s^{-1} for form I (Extended Data Fig. 5 and Extended Data Table 2). The unusual catalytic properties of MccA can be rationalized from its active site, a shallow cavity on the distal side of haem 2, lined by residues K208, Y285, Y301, R366 and K393, which are conserved among MccA orthologues (Fig. 3a). In the copper-depleted form II of MccA, the absence of the heterometal allowed for a binding mode of sulfite that was similar to the one seen in the sirohaem-containing enzymes or in NrfA (Extended Data Fig. 4)[4,21]. In the structure of the Cu-containing, high-activity form I of MccA, however, all 12 monomers in the asymmetric unit had a ligand bound to haem 2. While in most cases this was a mixture of two species, the electron density maps for some monomers were unambiguous, revealing one of two distinct species. In chain D this was an SO_2 molecule, coordinated through sulfur (Extended Data Fig. 6a). The two oxygen atoms of the ligand formed short hydrogen bonds to R366 and K208, respectively, with an additional interaction of both with K393 from the roof of the cavity. Furthermore, a water molecule contacted both sulfite oxygens and the hydroxyl groups of Y285 and Y301 (Fig. 3a). In chain H the dominant species was an SO molecule, probably generated through photoreduction during X-ray data collection or the exposure to H_2 in an N_2/H_2 atmosphere during oxygen-free crystallization. It coordinated the haem iron with its sulphur atom, which was readily distinguishable from the lighter oxygen (Extended Data Fig. 6b). The oxygen atom was hydrogen bonded to R366, and two water molecules were defined. One was bound to Y285 and Y301, while the other formed hydrogen bonds with K208 and Y285 (Fig. 3b). Overall, the hydrogen-bond network in the active-site cavity remained almost unchanged between both states, while a water molecule was removed from SO_2. No electron density was observed between the substrate molecule and copper, and the sulfur–copper distance was 3.2 Å for SO_2 and 3.3 Å for SO.

These snapshots of sulfite reduction at MccA suggest that the six-electron reduction to sulfide starts with the binding of sulfite to haem 2 and an immediate dehydration to $SO_2 \cdot H_2O$ (Figs 3a and 4a). The linear Cu(I)–S_2 arrangement supports this step with an overall negative charge, but the binding and dehydration does not yet involve reduction of the substrate and the SO_2 sulphur atom remains at the oxidation level S^{+IV} (Fig. 4a). The arrangement of haem groups in the protein then promotes three consecutive two-electron reduction steps, leading via $S^{+II} = O$ (Figs 3b and 4b), after elimination of a second water molecule, and S^0–O (Fig. 4c), where the third H_2O is released, to S^{-II} (Fig. 4d). Here, residue K393 plays an important part, as it forms hydrogen bonds to both oxygen atoms of SO_2, and also to the $O_η$ atom of Y285 (Fig. 3a). After reduction by two electrons, a water molecule moves towards Y285 by 1.5 Å, breaking its hydrogen bond to K393, but retaining the one to K208 (Figs 3b and 4b). The release of a third water molecule probably then severs the remaining hydrogen bond to K393 and possibly weakens the K393–Y285 interaction, leading to the deconstruction of the hydrogen-bond network in the active site and product release. The structures elucidate why MccA does not generate side products such as thiosulfate or trithionate through the closed hydrogen-bond network that holds all reaction partners tightly in place during catalysis and retains the oxygen atoms well defined as bound product water (Fig. 3b and Extended Data Fig. 6). The proposed mechanism does not require a change of oxidation state for the Cu(I) site, whose main task is to shift the dehydration equilibrium of the substrate towards SO_2. This view is supported by the observation that oxidation by air led to an irreversible loss of copper (Extended Data Fig. 4). In the presence of eight strongly absorbing haem groups, a single, transient Cu(II) may be difficult to detect (Extended Data Fig. 7a), but under no conditions did time-resolved ultraviolet/visible spectroscopy provide any indication of the occurrence of oxidized copper (Extended Data Fig. 7). Also, the sulfite substrate only oxidized half the haem groups of reduced MccA (Extended Data Fig. 8), emphasizing that the total electron charge of the multihaem enzyme is of high relevance for catalysis. Taken together, we have elucidated the structural and mechanistic basis of the most active bacterial sulfite reductase known to date, an enzyme that is assumed to contribute significantly to sulfite turnover in various habitats on Earth.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Simon, J. & Kroneck, P. M. H. Microbial sulfite respiration. *Adv. Microb. Physiol.* **62,** 45–117 (2013).
2. Kern, M., Klotz, M. G. & Simon, J. The *Wolinella succinogenes mcc* gene cluster encodes an unconventional respiratory sulphite reduction system. *Mol. Microbiol.* **82,** 1515–1530 (2011).
3. Hartshorne, R. S. *et al.* A dedicated haem lyase is required for the maturation of a novel bacterial cytochrome *c* with unconventional covalent haem binding. *Mol. Microbiol.* **64,** 1049–1060 (2007).

4. Lukat, P. *et al.* Binding and reduction of sulfite by cytochrome *c* nitrite reductase. *Biochemistry* **47,** 2080–2086 (2008).

5. Kemp, G. L. *et al.* Kinetic and thermodynamic resolution of the interactions between sulfite and the pentahaem cytochrome NrfA from *Escherichia coil*. *Biochem. J.* **431,** 73–80 (2010).

6. Pereira, I. A. C., LeGall, J., Xavier, A. V. & Teixeira, M. Characterization of a heme *c* nitrite reductase from a non-ammonifying microorganism, *Desulfovibrio vulgaris* Hildenborough. *Biochim. Biophys. Acta* **1481,** 119–130 (2000).

7. Steuber, J., Arendsen, A. F., Hagen, W. R. & Kroneck, P. M. H. Molecular properties of the dissimilatory sulfite reductase from *Desulfovibrio desulfuricans* Essex and comparison with the enzyme from *Desulfovibrio vulgaris* Hildenborough. *Eur. J. Biochem.* **233,** 873–879 (1995).

8. Shirodkar, S., Reed, S., Romine, M. & Saffarini, D. The octahaem SirA catalyses dissimilatory sulfite reduction in *Shewanella oneidensis* MR-1. *Environ. Microbiol.* **13,** 108–115 (2011).

9. Klotz, M. G. *et al.* Evolution of an octahaem cytochrome *c* protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria. *Environ. Microbiol.* **10,** 3150–3163 (2008).

10. Einsle, O. *et al.* Cytochrome *c* nitrite reductase from *Wolinella succinogenes*. Structure at 1.6 Å resolution, inhibitor binding, and heme-packing motifs. *J. Biol. Chem.* **275,** 39608–39616 (2000).

11. Iverson, T. M. *et al.* Heme packing motifs revealed by the crystal structure of the tetra-heme cytochrome c_{554} from *Nitrosomonas europaea*. *Nature Struct. Biol.* **5,** 1005–1012 (1998).

12. Einsle, O. Structure and function of formate-dependent cytochrome *c* nitrite reductase, NrfA. *Methods Enzymol.* **496,** 399–422 (2011).

13. Einsle, O. *et al.* Structure of cytochrome *c* nitrite reductase. *Nature* **400,** 476–480 (1999).

14. Igarashi, N., Moriyama, H., Fujiwara, T., Fukumori, Y. & Tanaka, N. The 2.8 Å structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium, *Nitrosomonas europaea*. *Nature Struct. Biol.* **4,** 276–284 (1997).

15. Polyakov, K. M. *et al.* High-resolution structural analysis of a novel octaheme cytochrome *c* nitrite reductase from the haloalkaliphilic bacterium *Thioalkalivibrio nitratireducens*. *J. Mol. Biol.* **389,** 846–862 (2009).

16. Mowat, C. G. *et al.* Octaheme tetrathionate reductase is a respiratory enzyme with novel heme ligation. *Nature Struct. Mol. Biol.* **11,** 1023–1024 (2004).

17. Hartshorne, S., Richardson, D. J. & Simon, J. Multiple haem lyase genes indicate substrate specificity in cytochrome *c* biogenesis. *Biochem. Soc. Trans.* **34,** 146–149 (2006).

18. Kern, M., Eisel, F., Scheithauer, J., Kranz, R. G. & Simon, J. Substrate specificity of three cytochrome *c* haem lyase isoenzymes from *Wolinella succinogenes*: unconventional haem *c* binding motifs are not sufficient for haem *c* attachment by Nrfl and CcsA1. *Mol. Microbiol.* **75,** 122–137 (2010).

19. Simon, J. & Hederstedt, L. Composition and function of cytochrome *c* biogenesis System II. *FEBS J.* **278,** 4179–4188 (2011).

20. Aoyama, H. *et al.* A peroxide bridge between Fe and Cu ions in the O_2 reduction site of fully oxidized cytochrome *c* oxidase could suppress the proton pump. *Proc. Natl Acad. Sci. USA* **106,** 2165–2169 (2009).

21. Schiffer, A. *et al.* Structure of the dissimilatory sulfite reductase from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *J. Mol. Biol.* **379,** 1063–1074 (2008).

Author Contributions B.H., M.K. and L.L.P. performed the experiments, B.H. and O.E. built and refined the structural model, all authors designed the experiments, B.H., O.E. and J.S. wrote the manuscript.

Author Information Structural models and structure factors have been deposited in the Protein Data Bank under accession numbers 4RKM (form I) and 4RKN (form II). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.E. (einsle@biochemie.uni-freiburg.de) or J.S. (simon@bio.tu-darmstadt.de).

METHODS

Protein purification and crystallization. MccA was produced from cells of mutant *W. succinogenes* MccA–2×Strep under fumarate-respiring conditions, as a variant with a C-terminal double Strep-tag II[2,22]. The cells were pelleted at 6,000*g*, resuspended in 50 mM Tris/HCl buffer, pH 8.0, and 150 mM NaCl, and disrupted by two passages through a French pressure cell at 138 MPa (SLM Aminco). The lysate was cleared by centrifugation at 100,000*g* for 60 min. The supernatant containing the target protein was loaded onto Streptactin affinity resin (IBA) and eluted in a single step with 2.5 mM of D-desthiobiotin. The eluted protein-containing fractions were further purified by size-exclusion chromatography on a HiLoad 26/60 column with Superdex 200 prep grade material (GE Healthcare) that was pre-equilibrated and run in 20 mM Tris/HCl, pH 7.5, 150 mM NaCl. MccA eluted in two peaks, representing the trimeric and monomeric forms of the protein, respectively. For anoxic preparations, dioxygen was removed from all buffers by eight cycles of vacuum/N_2 using modified Schlenk techniques. Protein was isolated using closed systems, and further handling was carried out in an anoxic chamber at <5 p.p.m. pO_2. For crystallization only the trimer fraction was used and concentrated to 6 mg ml^{-1} for MccA form I and to 10 mg ml^{-1} for form II by ultrafiltration in Vivaspin 20 concentrators (50 kDa molecular weight cut-off, Sartorius).

Crystallization and X-ray diffraction data collection. Well-diffracting crystals of both forms I and II of MccA were obtained exclusively in an anoxic chamber at room temperature, using the sitting-drop vapour diffusion method. For MccA form II, 0.5 μl of protein solution (10 mg ml^{-1}) was mixed with 0.5 μl of a reservoir solution containing 13–17% (w/v) of polyethylene glycol 3,350 and 100 mM CHES/NaOH buffer at pH 8.5–9.0 and equilibrated against 100 μl of the same reservoir. Crystal growth required either the addition of the reductant sodium dithionite or co-crystallization with 200 mM of sodium sulfite. MccA form I crystallized from a solution of 6 mg ml^{-1}, with a reservoir solution of 10–15% (w/v) of polyethylene glycol 5,000 monomethyl ether, 100 mM HEPES/NaOH buffer at pH 7.0–7.5, 5% tacsimate, pH 7.0 (Hampton Research) and 4 mM Zwittergent 3-12 (Anatrace) without addition of substrate or reaction intermediates. Diffraction data sets were collected on beamline X06SA at the Swiss Light Source (Villigen, Switzerland) using a Pilatus 6M detector (Dectris). Crystals of the copper-depleted MccA form II belonged to the rhombohedral space group *H*3, with unit cell dimensions of *a* = *b* = 186.5 Å and *c* = 232.8 Å and four monomers per asymmetric unit, while MccA form I crystallized in monoclinic *P*2₁, with *a* = 211.1 Å, *b* = 103.0 Å, *c* = 233.1 Å, and *β* = 98.5° and had 12 monomers per asymmetric unit. For phase determination, a data set of a form II crystal was collected at the Fe K-edge at $\lambda = 1.739$ Å. Diffraction data for the generation of anomalous double difference maps for identification of the additional metal site were collected at the Co K-edge at $\lambda = 1.608$ Å, the Ni K-edge at $\lambda = 1.488$ Å, the Cu K-edge at $\lambda = 1.381$ Å, the Zn K-edge at $\lambda = 1.283$ Å and the As-edge at $\lambda = 1.044$ Å. Diffraction data were indexed and integrated with XDS[23] and scaled with AIMLESS[24] from the CCP4 suite[25]. For data processing statistics see Extended Data Table 1.

Structure determination. Phase information for MccA form II was obtained by single-wavelength anomalous diffraction using data collected at the Fe K-edge. All 32 Fe sites and several sulphur positions were located with SHELXD[26] and refined with SHARP[27]. AutoSHARP was used for phase calculation, density modification and initial model building[28]. The structure was rebuilt and refined in iterative steps using COOT[29] and REFMAC5[30] to final R_{cryst}/R_{free} values of 0.16/0.18 at 2.1 Å resolution. This model was subsequently used to solve the structure of MccA form I by molecular replacement with MOLREP[31]. All structures were validated with MolProbity[32]. Refinement and validation statistics are summarized in Extended Data Table 1.

Metal quantitation. The Cu content of MccA was determined by ICP-MS (Spurenanalytisches Laboratorium Dr Baumann, Maxhütte-Haidhof, Germany) to 0.7–1.0 Cu per monomer and 0.2–0.5 Cu per monomer for different preparations of forms I and II, respectively. Owing to the interfering presence of sodium ions the error of the analysis was large, but the electron density features observed in the corresponding crystal structures were in accordance with the obtained values.

Activity assays. The specific activity of MccA for sulfite and nitrite reduction was determined at 303 K using benzyl viologen as an artificial electron donor. In the absence of dioxygen, benzyl viologen (0.4 mM) was reduced by oxalate (5 mM) via catalytic amounts of 5-deazaflavin upon illumination, and its reoxidation rate before and after addition of substrate was monitored at 562 nm. Final protein concentrations were 0.8 μg ml^{-1} for form I and 1.1 μg ml^{-1} for form II. The assay was conducted in 100 mM potassium phosphate buffer at pH 7.0. Substrate concentrations were assayed from 5 μM to 1 mM and kinetic parameters were determined from a sigmoidal fit (Extended Data Fig. 5).

22. Kern, M. & Simon, J. Production of recombinant multiheme cytochromes c in *Wolinella succinogenes*. *Methods Enzymol.* **486,** 429–446 (2011).
23. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
24. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69,** 1204–1214 (2013).
25. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67,** 235–242 (2011).
26. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
27. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59,** 2023–2030 (2003).
28. Vonrhein, C., Blanc, E., Roversi, P. & Bricoone, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
29. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
30. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
31. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66,** 22–25 (2010).
32. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
33. Cedervall, P., Hooper, A. B. & Wilmot, C. M. Structural studies of hydroxylamine oxidoreductase reveal a unique heme cofactor and a previously unidentified interaction partner. *Biochemistry* **52,** 6211–6218 (2013).
34. Weiss, M. & Hilgenfeld, R. On the use of the merging *R* factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **30,** 203–205 (1997).
35. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
36. Cruickshank, D. W. J. Remarks about protein structure precision. *Acta Crystallogr. D* **55,** 583–601 (1999).
37. Brünger, A. T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–475 (1992).
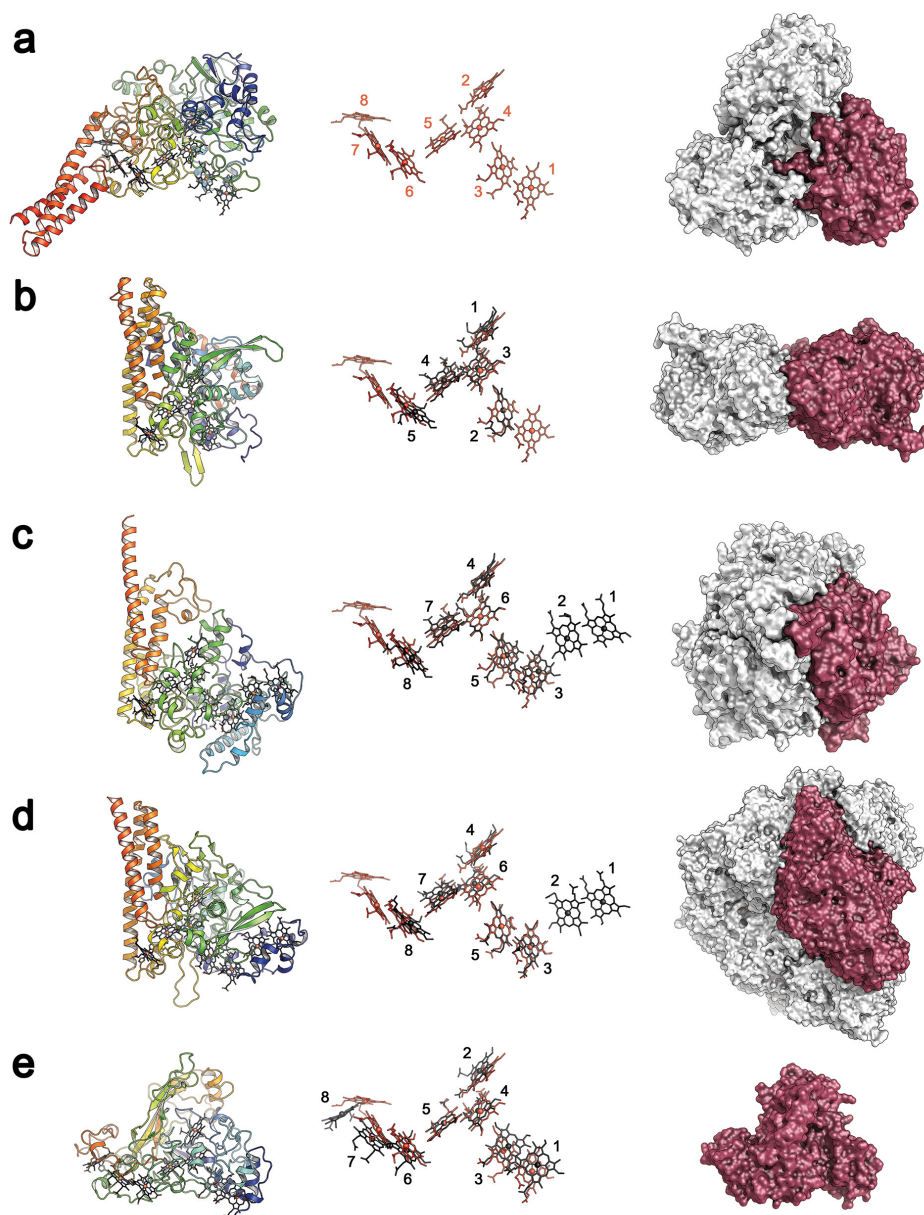
Extended Data Figure 1 | Comparison of monomer structure, haem packing and quaternary structure in multihaem *c* enzymes. a–e, Monomer structure is shown on the left, haem packing in the centre, and quaternary structure on the right. **a**, *W. succinogenes* sulfite reductase MccA. **b**, *W. succinogenes* nitrite reductase NrfA (PDB accession 1FS7)[10]. **c**, *N. europaea* hydroxylamine oxidoreductase (PDB accession 4FAS)[33]. **d**, Hexameric *T. nitratireducens* nitrite reductase (PDB accession 3SXQ)[15]. **e**, *S. oneidensis* tetrathionate reductase (PDB accession 1SP3). All haem arrangements are superimposed onto that of MccA (red).
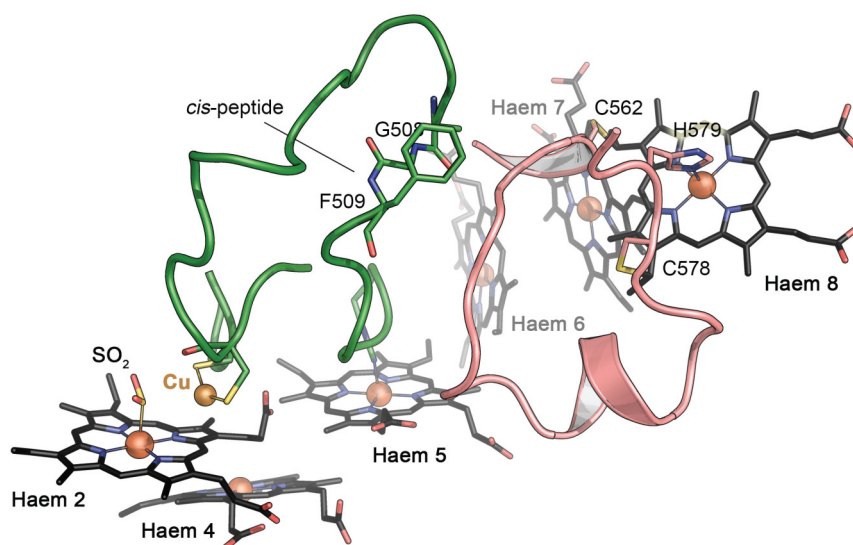
Extended Data Figure 2 | Uncommon structural features of MccA. Haem group 8 is linked to an unusual $CX_{15}CH$ motif that results in an extended loop region between C562 and C578 (red), dislocating haem 8 from the regular, perpendicular haem-packing pattern otherwise observed in multihaem enzymes. This region is adjacent to a loop starting at the copper ligand C495 that contains the only non-proline *cis*-peptide bond of MccA between G508 and F509. Conceivably, the action of the peptidyl isomerase MccB is required to generate this conformation after insertion of Cu(I), since the deletion of the *mccB* gene prevented the formation of stable MccA[2].
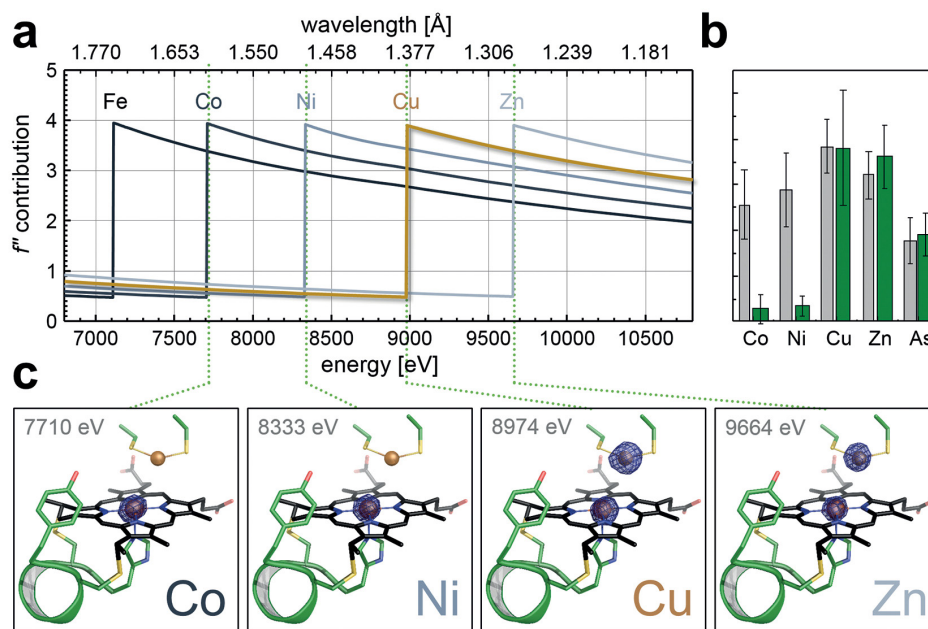
Extended Data Figure 3 | Identification of the additional heteroatom as copper based on anomalous difference electron density maps. a, Diffraction data sets were collected at the absorption edges of Co (7,710 eV), Ni (8,333 eV), Cu (8,974 eV), Zn (9,664 eV) and As (11,866 eV, not shown in plot). **b**, Integrated electron density is shown for a sphere with $r = 1.0$ Å around the Fe positions (grey) and the heterometal site (green) indicated in the individual data sets. The heterometal absorption properties behave as expected for Cu.

Error bars represent standard deviations for data from 12 individual observations in the asymmetric unit of the crystal form. **c**, Anomalous difference $(F^+ - F^-)$ electron density maps calculated for the individual data sets at the elemental absorption edges and contoured at the 5σ level. A maximum is observed only at energies corresponding to Cu and above, allowing for an unambiguous identification.
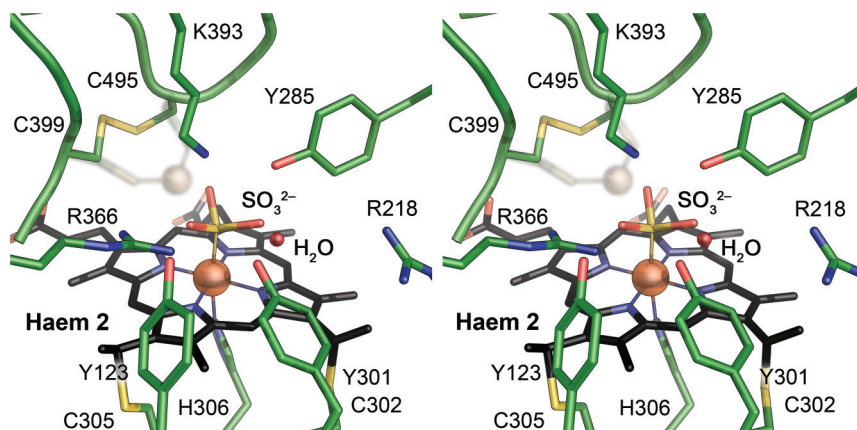
Extended Data Figure 4 | Structure of partially copper-depleted MccA. Protein isolated under oxic conditions commonly showed a decreased copper content of 0.2–0.5 Cu per monomer. The electron density surrounding the active site was best interpreted as a mixture of copper-containing monomers, probably with bound SO_2 as in Fig. 3a, and an oxidatively damaged, Cu-depleted form, where the thiol groups of C399 and C495 have rotated away from the active site to form a disulfide bond. This leads to improved access to the active site at the distal axial position of the iron atom of haem 2, allowing for the binding of SO_3^{2-}. The short distance of 1.4 Å between a sulfite oxygen and the Cu(I) ion indicated that the two indeed represent alternative and mutually exclusive conformations, forms I and II, that overlay in the crystal.
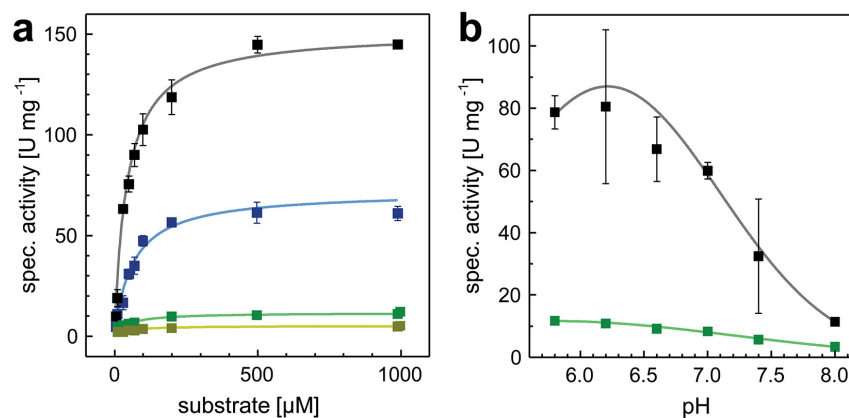
Extended Data Figure 5 | Representative steady-state kinetic analysis of MccA. **a**, The reactivity for sulfite and nitrite was assayed by following the reoxidation of benzyl viologen. MccA form I, isolated in the strict absence of dioxygen, showed higher activities for sulfite (black) and nitrite (green) than form II, isolated under oxic conditions (sulfite, blue; nitrite, yellow). For sulfite reduction, the kinetic parameters obtained were $v_{max} = 151.3 \pm 0.5 \, U \, mg^{-1}$ and $K_M = 43.9 \pm 2.3 \, \mu M$ for form I and $v_{max} = 71.8 \pm 0.5 \, U \, mg^{-1}$ and $K_M = 63.9 \pm 9.4 \, \mu M$ for form II (see Extended Data Table 2 for parameters). **b**, pH dependence of the sulfite-reducing (black) and nitrite-reducing (green) activity of MccA. Sulfite reduction shows an apparent maximum at pH 6.2. All measurements were carried out in triplicate with the error bars representing standard deviations between individual experiments.
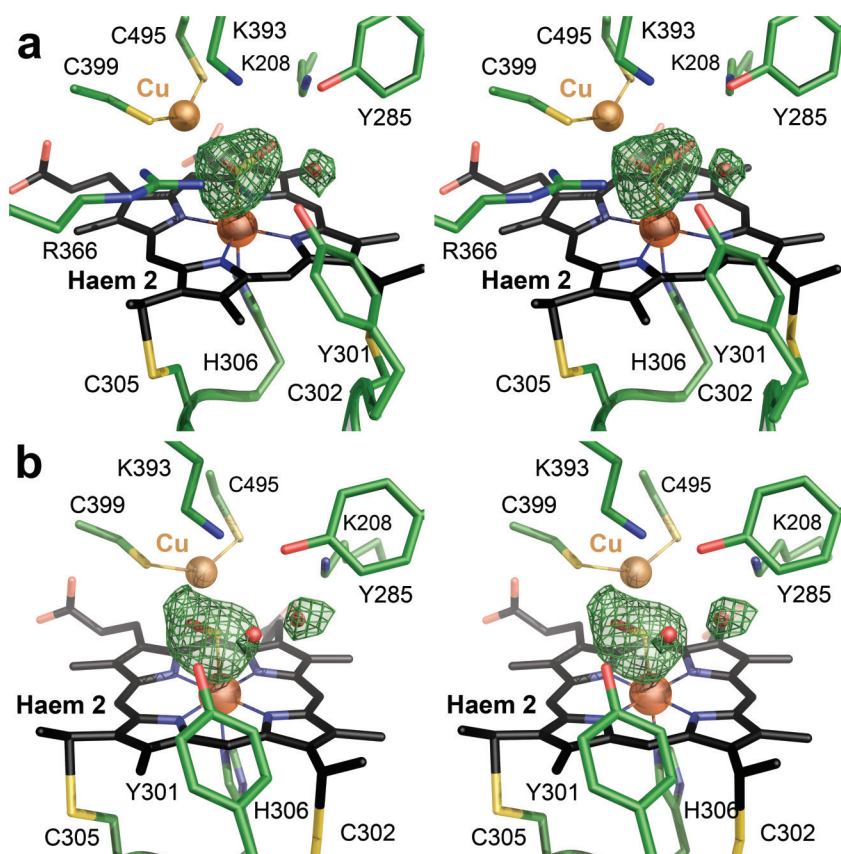
Extended Data Figure 6 | Difference electron density maps for bound substrate and intermediates. a, b, The stereo images show monomer D (**a**) and monomer H (**b**) of the $P2_1$ structure of form I MccA (Extended Data Table 1). Whereas in **a** an SO_2 molecule is bound to the Fe ion of haem group 2 together with a single H_2O molecule, photoreduction has led to the generation of a second H_2O, leaving SO bound to haem. The quality of the electron density maps allowed us to discern S and O atoms, but the state depicted in **b** may represent either an S(II) = O adduct (Fig. 4b) or the further reduced S(0)–O (Fig. 4c). All maps are $F_o − F_c$ omit difference electron density maps calculated in the absence of ligands.
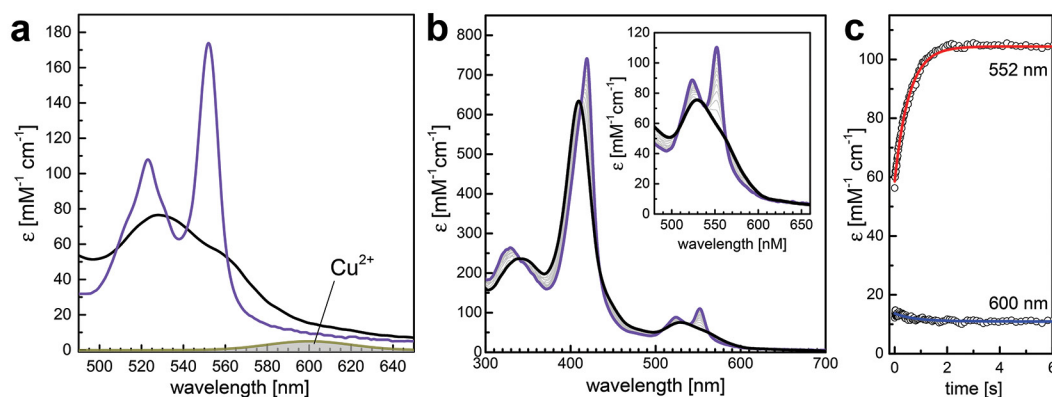
Extended Data Figure 7 | Copper detection by time-resolved electron excitation spectroscopy of *W. succinogenes* MccA. a, Owing to the presence of eight haem groups, a hypothetical, calculated absorption band of a typical type I copper site with an extinction coefficient of approximately $5,000\,M^{-1}\,cm^{-1}$ at 600 nm (grey) would be difficult to detect in a background of oxidized (black) or reduced (purple) haems, but may possibly be extracted from the well-characterized absorption of the porpyrin cofactors. **b,** Form I

MccA as isolated shows all haem groups in the oxidized state, with copper present as Cu(I) (black trace). Reduction with 100-fold molar excess of Ti(III) citrate in a stopped-flow spectrophotometer rapidly leads to the emergence of a typical ferrohaem spectrum (purple). For oxidation of reduced protein see Extended Data Fig. 8. **c,** Kinetic traces at 552 nm (red) and 600 nm (blue) show that the rise of the α-band at 552 nm is not accompanied by a relevant change at 600 nm.
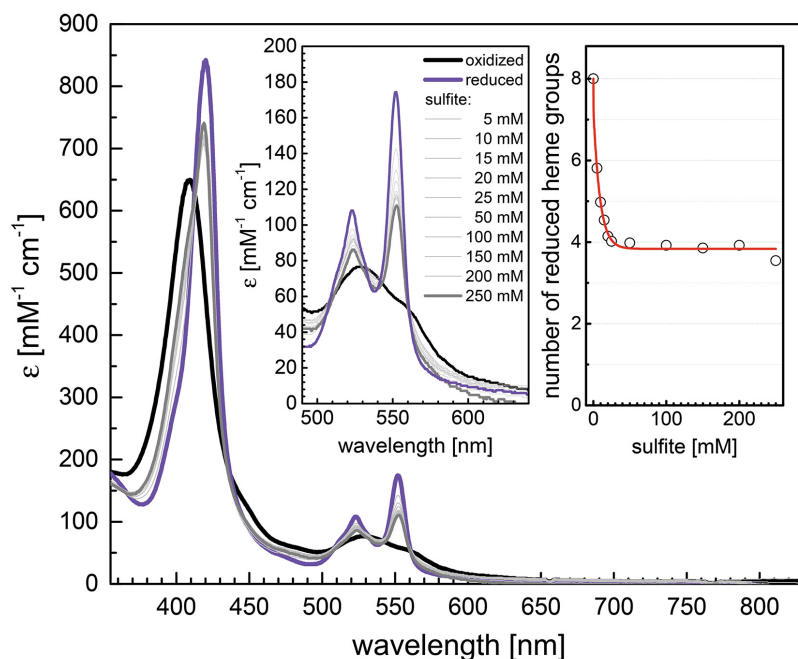
Extended Data Figure 8 | Reoxidation of *W. succinogenes* MccA with sulfite. MccA as isolated in the oxidized state (black) and subsequently fully reduced by Ti(III) (purple) to be reactive towards sulfite. After removal of excess reductant, the stepwise addition of sulfite to reduced MccA leads to successive oxidation of haem centres (inset left, grey spectra). However, even at a large excess of sulfite (250 mM) the oxidation of MccA is incomplete. Assuming nearly identical spectroscopic properties for all haem groups, the number of reduced haems can be deduced from the relative position of the α-band at 552 nm between the oxidized state (0 haems reduced, black) and the Ti(III)-reduced state (8 haems reduced, purple). This evaluation shows that sulfite only obtains up to 4 electrons from fully reduced MccA (inset right). As the enzyme forms a trimer with 24 haem groups, a total of 12 electrons are drawn from the system, sufficient to fully reduce 2 sulfite molecules to sulfide.

Extended Data Table 1 | Data collection and refinement statistics

Data Set	MccA form I (anoxic)	MccA form II (oxic)
PDB accession code	4RKM	4RKN
space group	$P2_1$	$H3$
cell constants a, b, c [Å]	211.1, 103.0, 233.1	186.5, 186.5, 232.8
α, β, γ [°]	90.0, 98.5, 90.0	90.0, 90.0, 120.0
resolution limits [Å]	49.37 – 2.20 (2.24 – 2.20)	47.22 – 2.10 (2.14 – 2.10)
completeness (%)	100 (99.9)	99.8 (96.6)
unique reflections	500,807 (24,733)	175,765 (8,426)
multiplicity (%)	6.6 (6.4)	9.4 (5.8)
R_{merge} *	0.199 (1.774)	0.102 (0.430)
$R_{p.i.m.}$ [34]	0.084 (0.758)	0.035 (0.188)
mean I/ σ(I)	9.2 (1.2)	17.3 (3.6)
$CC_{1/2}$ [35]	0.992 (0.506)	0.998 (0.843)
refinement statistics		
R_{cryst} †	0.18	0.16
R_{free} (%)	0.22	0.18
non-hydrogen atoms	70,014	24,154
solvent molecules	3,144	1,464
Cruickshank's DPI [36]	0.246	0.205
figure of merit	0.802	0.901
r.m.s. deviations from ideal values		
bond lengths (Å)	0.009	0.007
bond angles (°)	1.3	1.3
average B values (Å^2)		
protein main chain atoms	38.98	25.51
protein all atoms	39.98	26.03
solvent	35.65	30.43

References 34–36 are cited within the table. R_{free} is the cross-validation R value for a test set of 5% of unique reflections[37].

* $R_{merge} = \Sigma_{hkl} [(\Sigma_i | I_i - \langle I \rangle |)/\Sigma_i I_i]$

† $R_{cryst} = \Sigma_{hkl} || F_{obs} | - | F_{calc} || / \Sigma_{hkl} | F_{obs} |$

Extended Data Table 2 | Kinetic parameters for MccA

enzyme	substrate	v_{max}	K_M	k_{cat}	k_{cat}/K_M
form I (anoxic)	SO_3^{2-}	151.3±0.5	43.9±2.3	200.6	3.44
	NO_2^-	11.6±0.3	44.0±5.0	16.1	0.37
form II (oxic)	SO_3^{2-}	71.8±0.5	63.9±9.4	99.5	1.55
	NO_2^-	5.1±0.2	39.9±4.9	7.1	0.18

Kinetic parameters were recorded at 303 K. Units for v_{max} are $\mu mol\,min^{-1}\,mg^{-1}$; K_M is in μM; k_{cat} is in s^{-1}; and for k_{cat}/K_M units are $\mu M\,s^{-1}$.

CORRECTIONS & AMENDMENTS

Corrigendum: Derivation of novel human ground state naive pluripotent stem cells

Ohad Gafni, Leehee Weinberger, Abed AlFatah Mansour, Yair S. Manor, Elad Chomsky, Dalit Ben-Yosef, Yael Kalma, Sergey Viukov, Itay Maza, Asaf Zviran, Yoach Rais, Zohar Shipony, Zohar Mukamel, Vladislav Krupalnik, Mirie Zerbib, Shay Geula, Inbal Caspi, Dan Schneir, Tamar Shwartz, Shlomit Gilad, Daniela Amann-Zalcenstein, Sima Benjamin, Ido Amit, Amos Tanay, Rada Massarwa, Noa Novershtern & Jacob H. Hanna

The reduced representation bisulphite sequencing (RRBS) data generated and analysed in this Letter were not originally uploaded to the Gene Expression Omnibus (GEO), but can now be found under accession number GSE52617 (within the GSE52824 super series). As can be seen in this RRBS data, mouse naive pluripotent cells expanded for an extended period (that is, 90 days) in naive 2i/LIF (inhibition of ERK1/2 and GSK3β signalling plus leukaemia inhibitory factor)) conditions, globally lose DNA imprinting.

CORRECTIONS & AMENDMENTS

Corrigendum: Deterministic direct reprogramming of somatic cells to pluripotency

Yoach Rais, Asaf Zviran, Shay Geula, Ohad Gafni,
Elad Chomsky, Sergey Viukov, Abed AlFatah Mansour,
Inbal Caspi, Vladislav Krupalnik, Mirie Zerbib, Itay Maza,
Nofar Mor, Dror Baran, Leehee Weinberger, Diego A. Jaitin,
David Lara-Astiaso, Ronnie Blecher-Gonen, Zohar Shipony,
Zohar Mukamel, Tzachi Hagai, Shlomit Gilad,
Daniela Amann-Zalcenstein, Amos Tanay, Ido Amit,
Noa Novershtern & Jacob H. Hanna

In this Article, the reduced representation bisulphite sequencing (RRBS) data obtained and analysed were not initially uploaded to the Gene Expression Omnibus (GEO) record, but can now be found under accession number GSE64115, within the GSE49767 super series. Furthermore, gene expression data for two samples ($Mbd3^{+/+}$ embryonic stem cells and $Mbd3^{+/+}$ mouse embryonic fibroblasts (MEFs) after 8 days on doxycycline (dox)) were generated previously by our group[1] and deposited in the GEO under accession numbers GSM874650 and GSM874654. Further analysis of this data set was included in Figs 3a, 5c and Extended Data Fig. 8a, c of this Article. Also, in Fig. 5c and Extended Data Fig 8c of this Article, the '$Mbd3^{+/+}$ MEF + 8 day dox' sample was inadvertently mislabelled as '$Mbd3^{+/+}$ MEF + 11 day dox'. This change does not influence the conclusions of the manuscript.

In addition, further details on the experimental settings used to generate the data in the GSE49766 series reported in this Article (under the GSE49767 super series) have been added to the GEO website. It now states that in Fig. 3a and Extended Data Figs 5 and 8, out of 20 independent clonal series generated in our study carrying either the GOF18 ΔPE-Oct4-GFP transgenic reporter (Addgene plasmid 52382) or the complete GOF18 Oct4-GFP transgenic reporter (Addgene plasmid 60527), the clonal series selected for genomic analysis included an $Mbd3^{+/+}$ clone that carries the GOF18 ΔPE-Oct4-GFP transgenic reporter, and $Mbd3^{flox/-}$ and $Mbd3^{-/-}$ cells that carry the GOF18 Oct4-GFP transgenic reporter (complete *Oct4* enhancer region with distal and proximal enhancer elements). These reporters can be identified when analysing the *Oct4* locus in genomic DNA input datasets. As we do not use Oct4–GFP or any other selection for sorting cells before conducting genomic experiments, the difference in transgene reporters would not influence the interpretation of our genomic analysis data in any way. In these genomics studies, the endogenous *Nanog* and *Oct4* loci are not manipulated and are identical between all cell lines because the Oct4–GFP reporters were introduced via random transgenesis and validated for specificity. Notably, for the mouse induced pluripotent stem (iPS) cell efficiency results presented in our Article (for example, in Figs 1 and 2 and Supplementary Videos 1–4), cell lines carried the matched Nanog–GFP knock-in reporter or Oct4–GFP transgenic reporter (containing both distal and proximal enhancer elements as delineated in Extended Data Fig. 3a). Thus, all iPS cell efficiency and kinetic comparisons were conducted by using matched and validated pluripotency reporter systems.

1. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488,** 409–413 (2012).

CAREERS

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

Global-health specialist Adrian Hill examines a candidate Ebola vaccine for a ground-breaking clinical trial.

PUBLIC HEALTH

Behind a vaccine

A burgeoning vaccine industry paves paths for scientists to spread their knowledge outside the laboratory.

BY BRYN NELSON

In August 2014, officials from the World Health Organization (WHO) called global-health specialist Adrian Hill, who is the director of a non-profit vaccine-research centre. They had an urgent question: how soon could the centre launch a clinical trial for an Ebola vaccine?

"They weren't talking months — they were talking weeks, if not days," says Hill, who works at the Jenner Institute in Oxford, UK. A non-profit partnership between the University of Oxford and the animal-health-focused Pirbright Institute in Woking, UK, the Jenner Institute had vaccine-research programmes in progress that targeted nearly 20 human and veterinary diseases — but not Ebola.

Yet within a month of the phone call, the institute had launched an Ebola-research initiative. And six months later, it was testing candidate vaccines, including one in a phase III trial that involved 27,000 people in Liberia.

The rapid implementation of this programme underscores the swift pace and nimble nature of vaccinology, in sharp contrast to other disciplines in which it can take decades to gather results and create practical applications. And although priorities can shift rapidly, the pressing nature of outbreaks means that researchers can see their ideas implemented quickly to tackle a major disease. "There are not many fields where you can go in as a graduate student to a lab and finish up four years later, and something you've made with your own hands at the bench is being used to immunize people in a clinical trial that you are part of," Hill says.

Early-career scientists with an eye on vaccinology do not necessarily need a PhD. Because the field is geared towards translating research into practice (see 'Vaccination campaigns must deliver trust first'), many employers value practical lab experience and a humanitarian mindset, and are willing to provide on-the-job training to promising candidates.

Getting a vaccine from the lab to the clinic requires an approach that includes fields from microbiology to chemical engineering. "One of the attractions for a young person coming into this field is that it's so diverse," says Hill. "You can head off in different directions." Opportunities exist throughout industry, the non-profit sector and academia.

VACCINATION CONTAGION

The global vaccine market has boomed in the past couple decades. Between 2000 and 2013, the market value for vaccines soared from US$5 billion to almost $24 billion; by 2025, that value is expected to quadruple. Groups around the world are trying to devise effective vaccines for dozens of diseases, especially the 'big three' — HIV/AIDS, malaria and tuberculosis.

In 2000, the Bill & Melinda Gates Foundation in Seattle, Washington, began pouring money into the development and distribution of vaccines for economically disadvantaged nations. After 11 September 2001, the US and other governments began to prioritize vaccine research for diseases that could be used in bioterrorism. And the WHO and other aid organizations have helped to drive home the necessity of controlling diseases in developing countries, such as the ongoing Ebola outbreak in West Africa.

Novavax, a vaccine research-and-development company in Gaithersburg, Maryland, has more than tripled in size over the past 4 years to roughly 300 employees. "Novavax is hiring like crazy," says Christi McDowell-Patterson, director of upstream process development for the company. Her department, a mix of chemical engineers, cell biologists and other scientists, manages the cell lines and equipment used to create the company's vaccine candidates, including ones for Ebola, influenza and respiratory syncytial virus.

Novavax also hires summer interns — mainly undergraduates or recent graduates with bachelor's degrees. "If we like them, we'll try to figure out a way to keep them on," McDowell-Patterson says. Novavax, along with other companies and organizations — such as Sanofi, Novartis and the International AIDS Vaccine Initiative — offers tuition reimbursement to support graduate education for promising young scientists.

Apart from the many perks Novavax offers, ranging from yoga classes to a bowling league, chemist Natalie Thompson was most intrigued by the company's production methods, which use virus-like particles and nanoparticles composed of recombinant proteins. She joined the company in March 2014 after completing a PhD in analytical chemistry and a three-year postdoc during which she used mass spectrometry to study monoclonal antibodies.

At Novavax, she employs analytical methods such as liquid chromatography and mass spectrometry to determine the constituents and quality of the company's vaccine products. "What I enjoy about working in industry is that the final goal is really defined," she says. She also likes the fast pace, and that many different teams work together.

A SYMBIOTIC RELATIONSHIP

Many vaccine-development organizations have set up shop near universities to take advantage of local academic expertise, which means that opportunities for budding vaccinologists are often steps away. Gerald Strohmaier, global head of industrial relations and corporate finance at the biotechnology company Valneva, regularly draws talent from the University of Vienna, which hosts the company's main research facility. "We do not have a huge scouting department looking for talent all over the world," he says. "But whatever we can do with students, we do." Although Valneva hires only a few postdocs at a time, scientists who score a position have excellent job prospects — as many as 80% of them secure permanent jobs with the company.

Whereas pharmaceutical and biotech companies focus on more-marketable vaccines, many non-profit institutes are researching vaccines that have less commercial value. Salaries at these institutes are generally lower than those for university or industry jobs, but people do have the potential to make a humanitarian difference in developing countries.

The International Vaccine Institute (IVI), for example, is an independent organization at Seoul National University in South Korea that creates and introduces vaccines for neglected infectious diseases. One of the institute's most-successful developments is a low-cost oral vaccine for cholera. But that vaccine requires two doses given two weeks apart, and IVI researchers are in the process of analysing data from a clinical trial of 200,000 people in Bangladesh to gauge the effectiveness of a

WORDS BEFORE NEEDLES
Vaccination campaigns must deliver trust first

Partnerships with health centres in developing countries help to immunize local children.

Vaccine hesitancy — the reluctance to receive or engage in vaccination — is insidious and alive. In northern India, a polio-vaccination campaign had to dispel rumours about potential side effects. The rumours had been spread by parents who had objected to strangers vaccinating their children. In Kenya, religious leaders who had not been consulted before a tetanus-vaccine initiative revived a 20-year-old rumour that the vaccine would sterilize children.

Heidi Larson has seen countless permutations of this issue. So accustomed was she to dealing with it during her work at the United Nations children's charity UNICEF in New York City that she became known as the director of UNICEF's fire department. The organization sent her around the world to extinguish the fires of distrust and suspicion that often impede the delivery and acceptance of desperately needed vaccines.

An anthropologist by training, she had worked extensively on HIV and AIDS issues affecting adolescents in Nepal and the South Pacific. "I really had thought coming from AIDS into vaccines would be a bit bland," she says. "To my surprise, this whole area of the introduction of new vaccines saw more emotional and political challenge than I ever could have imagined."

Director of the Vaccine Confidence Project at the London School of Hygiene and Tropical Medicine, Larson is mapping the social, cultural and political challenges to vaccination efforts around the world, from northern Nigeria to southern California.

Skills in vaccine introduction, risk management and communication are often overlooked, but groups such as UNICEF and the World Health Organization are increasingly recognizing their value. As a result, career opportunities are emerging

for good communicators with a desire to combine science and public service, helping vaccination teams to earn the support of local communities. "I have more and more people contacting me who want to do masters and PhDs with our group and, at the same time, I have more and more demand for this type of work," Larson says. Some organizations are so desperate for help that Larson is developing a training course that can be completed more quickly than a masters or PhD.

Scientists can also do a graduate programme in public-health communication, and the US National Public Health Information Coalition and the Virtual Immunization Communication Network have compiled tools and training options.

Larson is leading the communication and preparedness effort for an Ebola-vaccine trial in Africa that will involve an initial shot and follow-up boost. She and her colleagues are engaging community leaders to work out acceptable procedures, from how to seek informed consent from the local population to how to distribute and administer the vaccine.

Even small misunderstandings can have profound consequences. The head of one vaccine programme related how ignorance of a local custom — addressing religious leaders and elders with honorifics to show respect — probably contributed to a year-long delay in getting approval for a mass typhoid-vaccination campaign in south Asia. When a colleague corrected the misstep, approval was gained in a few weeks.

For a vaccine trial in Gambia, one of Larson's graduate students modified the informed-consent process to align with the region's oral tradition, replacing written documents with an interactive video in the local language. The approach enhanced comprehension and retention, especially for people who had low levels of literacy.

"There's a real science behind how you introduce a new vaccine. It just doesn't happen by accident," says Peter Jay Hotez, president of the Sabin Vaccine Institute and Texas Children's Hospital Center for Vaccine Development in Houston. Beyond building partnerships, it also requires engaging in public dialogue.

Despite the dearth of formal programmes focused on vaccine outreach and communication, Hotez and other researchers say that those who have the skills and diplomacy can be crucial assets for ensuring that a vaccine-development scheme does not fall short of its goal. **B.N.**

NYANI QUARMYNE/PANOS

one-dose version to improve compliance. The institute is always on the lookout for research talent, says Sushant Sahastrabuddhe, a physician who heads the institute's Enteric and Diarrheal Disease Programme. Its international scientific staff of about 50 includes researchers with PhDs, medical doctorates and master's degrees.

The formation of partnerships between non-profit organizations and developing countries is also creating job opportunities. The Infectious Disease Research Institute (IDRI) in Seattle, Washington, employs about 125 people and works with nearly 100 collaborators around the world. It has helped to establish vaccine-formulation centres in South Africa and India. Manufacturing vaccines locally instead of importing them can decrease both the cost and distrust of the product, says Steven Reed, IDRI's founder.

The need to bolster public-health infrastructure in developing regions has become increasingly important. "There's minimal, almost zero, capacity to develop vaccines in the Middle East, north Africa and sub-Saharan Africa," says Peter Jay Hotez, president of the non-profit Sabin Vaccine Institute and Texas Children's Hospital Center for Vaccine Development in Houston. "And yet these are the places where we're going to see the next generation of catastrophic emerging infections." The US Department of State named Hotez as a US Science Envoy, a role designed to help promote international partnerships. For his project, he is pushing to expand vaccine infrastructure in Africa and the Middle East by focusing on countries such as Saudi Arabia and Morocco, which have an established scientific culture and PhD-level scientists. If he succeeds, programmes that increase vaccine production in these places could lead to more jobs for local scientists, as well as for international consultants.

> *The pressing nature of outbreaks means that researchers can see their ideas implemented quickly.*

For him, vaccinology provides the "perfect confluence" of humanitarian values and biomedical and social science. And as Hill can attest, the work is seldom boring. The past six months have been particularly "invigorating and energizing", he says, with regular enquiries about the Ebola trials from the British Prime Minister's office and other top government officials. "Going to the highest level of government with what you assayed yesterday," he says, "is pretty exciting." ■

Bryn Nelson *is a freelance writer based in Seattle, Washington.*

Metrics for ethics

Focus on perceived working conditions could help graduate schools to train responsible researchers.

BY MONYA BAKER

Training in research ethics is mandatory for many US graduate students and postdocs, but there is little evidence that formal classes prompt scientists to conduct research ethically. However, the workplace climate — which includes perceptions of regulatory committees, data confidentiality and treatment of trainees — influences research practices and can spawn behaviours such as poor record-keeping or plagiarism.

An interdisciplinary team has developed a survey to assess work conditions in research institutions, with a long-term goal of establishing a baseline for measurements of workplace climate across disciplines and universities. The SOuRCe (Survey of Organizational Research Climate) is a 32-question survey that divides workplace climate into seven categories, including integrity norms (such as giving due credit to others' ideas), integrity inhibitors (such as inadequate access to material resources) and adviser–advisee relations. The team hopes that such data will help institutions to craft policies that will improve research conduct.

The survey illuminates differences in attitudes held by faculty members and trainees, says Karen Klomparens, dean of the graduate school at Michigan State University (MSU) in East Lansing. When the school ran versions of the survey in 2009 and 2014, clusters of respondents reported feeling ill-equipped to judge whether university policies support responsible research — which suggests that those topics are not discussed in meaningful ways, she says. Klomparens used the results to spur faculty members in specific departments to talk to trainees about norms in authorship, data management and peer review. "Because we use the survey data by graduate programme and by discipline, we can make recommendations," she says. To encourage participation, she emphasized to respondents that the tool is not intended to shame or punish, and responses are stripped of identifying information.

Brian Martinson studies research integrity at the non-profit HealthPartners Institute for Education and Research in Bloomington, Minnesota, and helped to develop the survey at 40 academic health centres (B. C. Martinson *et al. Sci. Eng. Ethics* **19,** 813–834; 2013). He has also worked on it in a separate project with MSU, Pennsylvania State University and the University of Wisconsin–Madison. A poor workplace climate correlates with many undesirable research behaviours, even extreme forms such as data falsification, he explains (see 'Ethics in the environment'). Still, he thinks that institutions could boost integrity most effectively by focusing on common, less-attention-grabbing behaviours that are tightly tied to workplace climate, such as sloppy record-keeping. "They lower the standards that people are following over the long run," he says.

The dearth of robust, real-world research has long hampered efforts to improve integrity, says C. K. Gunsalus, director of the National Center for Professional and Research Ethics in Urbana, Illinois, who is working with Martinson's team to distribute the study online. "The climate survey provides actual empirical data," she says.

Gunsalus and others aim to aggregate results in a central database, so that a physics department at one institution, for instance, will be able to compare its climate scores with those of similar departments elsewhere.

Analysing workplace climate could be a powerful way to promote integrity, says Nicholas Steneck, a consultant for the US Office of Research Integrity in Rockville, Maryland, although he fears that institutions might find it tough to apply SOuRCe results to policy. But Gunsalus thinks that broad, quantitative survey results could make the issue more difficult to neglect. "The best thing that gets traction with scholars and scientists," she says, "is data." ■

ETHICS IN THE ENVIRONMENT

The level of self-reported fraud, fabrication and plagiarism increases as perceptions of 'integrity norms' fall and of 'integrity inhibitors' rise.
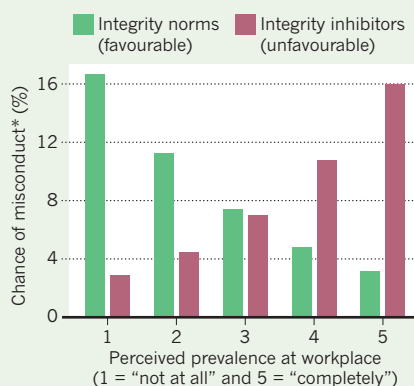
*Predicted likelihood that an individual would report engaging in fraud, fabrication or plagiarism, according to regression analysis of a large survey on research behaviour and workplace climate during the past three years.

THE ARCHIVE PERSONALITY PROTOCOL

A great way to resolve a problem.

BY BRIAN TRENT

"You can't *possibly* be drunk, Cynthia! You've been out of cryosleep for less than two hours!"

Cynthia Szabo reluctantly turned her gaze from the ship's viewscreen to regard her inquisitor. "Did you say something?"

Liza Neary's lips curled into a snarl and she was about to retort when the viewscreen flared again. The colony fleet — 20 massive Stanford torus vessels launched from Earth 300 years earlier — spread and contracted like a shoal of fish. Little explosions lit the void between them; a macabre light show from humanity's grand attempt at galactic diaspora.

"You *look* drunk," Liza breathed, bits of cryofrost still dangling from her eyebrows and dreadlocks. "Great. Just great, Cynthia."

"I'm *not* drunk," Cynthia slurred. "I'm merely rattled, okay? Three centuries of cryosleep, and *this* is what I wake up to!"

Liza laughed bitterly and squeezed the ends of her dreads. "We knew that the descendants of the original colonists would inevitably form their own societies. Knew that *theoretically* that might result in conflict, but …"

"How bad is it?"

"Want the laundry list?"

Cynthia nodded reluctantly.

"The *Thebes* went communist," Liza said, pointing to one of the many silver flecks on the screen. "They've sent a boarding party against the *Sparta*, which, while we were sleeping, became a loose confederation of libertarian enclaves. The fighting is a messy, level-by-level grudge match now."

"Why isn't the *Castor* helping? They're the flagship of the entire fleet …"

"Sure, and as the flagship, the *Castor* was settled by Earth's major entrepreneurial dynasties. Well *take a look.*" Liza tapped the screen and a real-time video bubble opened, displaying a feed of the *Castor*'s agora deck. Sweaty, wearied labourers moved in endless lines, hauling materials through a colour-coded maze of aisles, while uniformed officers monitored them from a cage-walk above.

Liza explained: "In 300 years, the *Castor*'s plutocratic elite formed monopolies on air, water and food supplies, so now they leverage their 'customer base' to gain a fraction of stock advantage against their corporate enemies."

Cynthia burped noisily, earning a scathing glare from her co-councillor. *Three centuries!* she thought. Seemed like only yesterday that she was climbing into her cryocapsule with the rest of the Emergency Protocol Council,

Earth still visible through the porthole. She had expected to sleep the whole way, the cryosleep suspension gas sinking her into a grey and restful slumber until …

She shook herself free of the memory. "Liza, you're not going to tell me the *entire fleet* has devolved into a free-for-all?"

Liza rubbed her chin. "No, that wouldn't be accurate. We've got two ships that reverted to anarcho-primitivism, with genome clans decrying all technology and raising their kids to believe that their seedships are the entire world. Then there's the *Pollux* …"

She opened another video bubble. A mass wedding ceremony was taking place. But something was wrong with all the *faces* …

"One eye is two inches higher than the other," Cynthia gasped. "What the hell?"

"I've been calling them the cubists. Their ancestors apparently decided that asymmetry is the epitome of beauty, so they implemented a merciless breeding programme. The result, over the past three centuries, has been these charming little Picasso people. Anyone born with symmetrical features is ground into mulch."

Cynthia felt sick.

"The rest of the fleet collapsed into oligarchies, despotisms, not one but *two* feudal states that make sixteenth-century Japan look like Disneyworld, four theocracies, a sopranocracy —"

"Should I even ask?"

"Rule by those with the highest-pitch voices. Sounds like a ship-wide helium leak."

Cynthia couldn't take it anymore. She pulled the hidden flask of vodka from her cryosuit pocket and took a mighty swig.

Liza's eyes grew wide. "You *are* drunk! That explains it!"

"Explains what? That humanity's grand attempt at colonizing the stars has become this free-for-all?"

"Thawed for two hours, and you still found time to get blitzed. Just great, Cynthia."

"Hey, at least I *did* something! The moment I realized the fleet had degenerated into chaos, I enacted the Archive Personality Protocol!"

Liza started to reply when suddenly two of the ships engaged in hostilities ceased firing. Another pair of ships followed suit.

"As I was saying," Cynthia continued, "the protocol seemed like the only option. It's —"

"A storage of reconstructed personalities, to help us deal with any possible contingency when we reach our destination," Liza snapped. "I know. Benjamin Franklin, Confucius, Saladin, Catherine, Marcus Aurelius …"

"I decided to use it now, to restore order to the fleet." She trailed off, as more and more ships began to converge into a tight, disciplined formation. The colourful bursts of weapon fire were winking out.

"Ha!" Cynthia exclaimed, the gleam of triumph in her inebriated eyes. "It worked! Good ol' President Hamilton was able to talk sense into everyone!"

Liza stared at her. "President Hamilton?"

"I figured he was the most innocuous, effective choice. He helped to stitch the divided American colonies into a cohesive nation, so …"

Liza checked the archive list of reconstructed personalities. "Alexander Hamilton?" she repeated.

"Yes, goddam it."

"You *were* drunk," Liza said, shaking her head. "You probably *meant* to conjure Mr Hamilton, but it looks like your finger slipped and selected the name above his."

Cynthia felt her smile melt. At that moment, a voice erupted from the ceiling speaker. It spoke in a language she didn't recognize.

Liza hit the TRANSLATE option, just in time to catch the tail-end of the message:

"— have failed to lead yourselves, and now I shall lead you. We shall conquer the very stars, world by world, and none shall oppose our unified might! Under my rule, we shall remake the very galaxy! I, Alexander, *swear this to you!*"

"Alexander?" Cynthia muttered. "The —"

"Great." Liza snatched the flask from her hand and downed a mighty swig. "Just great, Cynthia." ∎

Brian Trent *is a science-fiction writer and traveller. He lives in New England, somewhere between a city and a forest.*

ILLUSTRATION BY JACEY

nature**OUTLOOK**

ASSESSING SCIENCE

China's research under the magnifying glass

nature**OUTLOOK**

ASSESSING SCIENCE
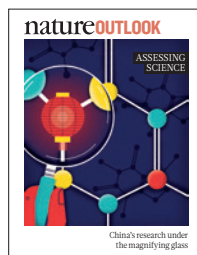
Cover art: Dale Edwin Murray

Over the past few decades, China has enjoyed economic prosperity, driven by its vigorous manufacturing output. More recently, the country has used this wealth to ratchet up other skills and now has the world's second largest spend on research and development (see page S8), and the second largest output of scientific papers.

Behind these accomplishments are the incentives that drive good scientific behaviour. Having lost a generation of scientists during the Cultural Revolution, the country has made a huge effort to instil qualities that are in line with those of other developed countries. These issues were examined at a symposium held in Shanghai in October 2014, co-hosted by *Nature*. Representatives from academia, industry and scientific societies talked about the importance of research to China's ongoing development, and how to stimulate it (S10). This Outlook was influenced by the debates, although *Nature* takes sole responsibility for all editorial content.

Unlike countries such as the United Kingdom and Australia, China does not have a national process of research assessment. The closest comparison is the evaluation undertaken by its largest research organization, the Chinese Academy of Sciences (S18).

Any complete measure of science must also recognize new technology, improved health and job creation — but Chinese researchers face institutional obstacles in translating their research into industrial applications (S28).

At the heart of this progress are China's scientists, yet many of them are struggling to find their way in a system that is undergoing such extensive change. In particular, young researchers (S36) and female scientists (S26) crave policies that would enable them to make a greater contribution. Such improvements may fortify China's scientific acumen and make the country a formidable player in global research.

Michelle Grayson
Senior Editor, Supplements

CONTENTS

RELATED ARTICLES
The latest news and comment on science in China published in *Nature*.

NANJING UNIVERSITY

A CENTURY-OLD CHINESE UNIVERSITY ASPIRING FOR EXCELLENCE

Founded in 1902, Nanjing University (NJU) is over a century old. It enjoys an excellent reputation in China and abroad. The historical development of the university has been closely connected with the mission of the nation and the needs of society. NJU's vision is to develop into a world-class, research-intensive and internationalized higher-education institution oriented toward high-calibre faculty, top-notch students and cutting-edge research.

Its president is Jun Chen, a professor in geochemistry and an academician of the Chinese Academy of Sciences. Presently, NJU has over 33,205 full-time students, of whom about 14,188 are undergraduate students, 15,667 are master and doctoral students and 3,350 are international students. NJU has three beautiful campuses: Gulou, Xianlin and Pukou. It has 28 schools containing 73 departments and has 8 national key disciplines in the primary disciplinary categories, 13 in secondary categories and 6 in the nurturing list. The university is listed in the first group of high-level universities in the national 211 and 985 projects, which attract strong funding from the Chinese government. As a member of the China Nine League, NJU boasts strong faculty in the humanities, social sciences, natural sciences, medicine and engineering.

Faculty

In its over 100 years' history, NJU has cultivated a great number of prominent and learned figures, and has consequently greatly contributed to the nation's revitalization and development. Many famous scientists and scholars have studied or worked at the university. Among its 2,251 faculty members, 28 are academicians of the Chinese Academy of Sciences, 3 are academicians of the Chinese Academy of Engineering, 1 is a foreign academician of the Chinese Academy of Sciences, 4 are academicians of the World Academy of Sciences, 1 is an academician of the Russian Academy of Sciences and 1 is an academician of the Academy of Science of the Royal Society of Canada. In addition, there are 115 Chair Professors and Distinguished Visiting Professors of the Chang Jiang Scholars Program, 97 winners of the National Distinguished Young Scientists Award, 29 overseas scholars under the National Thousand Talents Plan and 45 awardees of the National Youth Thousand Talents Plan.

Research

NJU is a leading comprehensive university in China as assessed by various national and international rankings. According to Thomson Reuters' Essential Science Indicators, 13 academic disciplines at NJU are ranked in the top 1 per cent worldwide and are ranked top among Chinese universities. In the QS Word University Rankings 2014, NJU has 20 disciplines listed in the global top 200.

Over the past decade, NJU has acquired more than 800 national, provincial and ministerial awards for research. Since 2000, NJU has won 1 first prize and 18 second prizes in the National Natural Science Awards, 6 second prizes in the National Scientific and Technological Progress Awards and 4 second prizes in the State Technological Awards. In 2006, a project accomplished by NJU physicists led by academician Nai-ben Ming won first prize in the National Natural Science Awards, which is the highest award for natural scientific research in China.

In recent years, NJU has undertaken dozens of state key projects. In addition, NJU ranks top among Chinese universities in terms of the quality and quantity of research papers listed in the Science Citation Index (SCI). The number of SCI publications of NJU was ranked first among Chinese universities for seven years since 1992, and NJU had the highest number of citations of SCI papers for eight years since 1992. In 2013, NJU was ranked fourth among Chinese universities in the Nature Publishing Index. In 2014, NJU was ranked 38th in the Global Nature Index and fourth for China. In 2015, NJU was ranked 7th in the Asia-Pacific Nature Index, and again fourth for China.

In addition to conducting basic scientific research, NJU actively pursues applied scientific and engineering research to promote economic and social development. NJU has established 23 platforms that involve cooperation among government, industry and university, to boost industry–academia interaction and the local economy. These platforms are for areas such as new materials, electronic data, biopharmaceuticals, environment and resources and fine chemicals.

International cooperation

NJU is very active in international cooperation and academic exchanges. It has
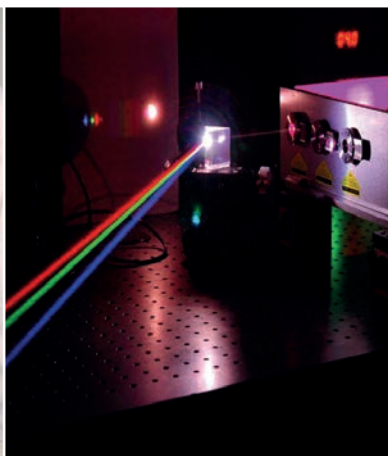
established partnership with over 280 institutions of higher learning and research institutes from over 30 different countries. NJU continually pursues opportunities for international cooperation. Such efforts resulted in the establishment of the Center for Chinese and American Studies with Johns Hopkins University, the Sino–German Institute for Law Studies with the University of Göttingen and the Center for Chinese and Japanese Studies with the University of Tokyo. Many world-renowned scholars, including Nobel laureates, are honorary or visiting professors of NJU. In 2011, Aaron Ciechanover, who was awarded the Nobel Prize for chemistry in 2004, accepted an invitation to serve as Director for the Institute of Chemistry and Biomedical Sciences of NJU. In 2013, Jean-Marie Gustave Le Clézio, who was awarded the Nobel Prize for literature in 2008, began to work at NJU as Distinguished Resident Scholar of the Institute of Advanced Studies in Humanities and Social Sciences of NJU.

Research highlights

Physical sciences

A briefing in *Nature* (**389,** 113–117, 1997) assessed the Laboratory of Solid State Microstructures of NJU as "approaching world-class standards in research." Of the numerous achievements at NJU, perhaps the most influential was that by Nai-ben Ming on the dielectric superlattice (DSL). For this work, he and his team members were awarded first prize in the National Natural Science Awards in 2006.

The team investigated the fundamental theory of domain engineering and quasi-phase-matching and applied it in the design, fabrication and application

of DSL materials. In recent years, Shi-ning Zhu's group successfully developed and commercialized a series of DSL based solid-state lasers, including mid-infrared tunable lasers, which have important applications in non-invasive medical diagnostics, free-space communication, laser scalpels, spectroscopy and remote sensing. Zhu's team has also done cutting-edge work in the fields of plasmonics and integrated photonic circuits. In particular, Hui Liu and co-workers demonstrated mimicking of gravitational lensing and optical trapping in an optical chip.

Excited by the striking properties and applications of DSLs in nonlinear optics, Zhu and Xu focused on the generation and manipulation of photonic entanglement — a key phenomenon in quantum technologies. The unique properties of DSLs enable the photon flux and achievable wavelength to be increased and also offer a compact way to tailor the polarization, frequency and spatial entanglement.

The team fabricated the first lithium niobate (LN) quantum chip that had efficient two-photon sources and manipulated photons using components such as junctions and wavelength-selective filters. This chip is controlled by on-chip electro-optical modulators, resulting in a variety of quantum states. "An LN superlattice may become an ideal platform for realizing fully integrated quantum optics," explains Ping Xu.

Chemistry

Nanomaterials and nanostructures are important in many fields. As one of the earliest nanoscience groups in China, Hong-Yuan Chen and Jing-Juan Xu's group in the State

Key Laboratory of Analytical Chemistry for Life Sciences has introduced nanotechnology and biotechnology to the field of electrochemistry and has conducted substantial experimental and theoretical studies in ultramicroelectrodes, electrochemical biosensing, bioelectrochemistry as well as microfluidic chips toward various life-related substances. Their work about the basic study of microelectrodes has been included in the official "Technical Report" by IUPAC in 2000. Interested in a broad range of functionalized nanomaterials and patterned nanostructures, they have exploited the unique electronic, mechanical, electrochemical and interfacial properties of these materials to address critical issues in both fundamental studies and real-world applications toward various life-related substances for more than 20 years. Their accomplishments have been recognized by national-level scientific awards.

Their current research focuses on the in-depth study of the electrochemical and photophysical properties of nanomaterials for the development of innovative bioanalytical techniques based on their photoelectric interconversion processes, i.e., electrochemiluminescent (ECL) and photoelectrochemical (PEC) bioanalysis. They were the first to combine the energy transfer with the ECL of semiconductor nanocrystals system and proposed the dual-potential ECL ratiometric sensing approach as well as the wireless bipolar electrode in bioanalysis. By means of the energy transfer processes between CdS nanocrystals and noble metal nanoparticles in PEC sensing systems, they further realized the bioanalysis with high sensitivity and specificity.

Earth sciences

Atmospheric dust plays an important role in the marine and terrestrial geochemical cycles and impacts the global climate on daily and orbital (10^4–10^5 years) timescales. Asia is the second largest source of dust on Earth; the deserts and sand fields in China dominate Asian dust emission. Studies of dust transport based on dust tracers and satellite imagery clearly show that aeolian dust from Asia is transported globally. In the leeward regions of dust source areas, dust deposition provides an excellent archive for reconstructing the past environmental changes since at least 22 million years ago.

Jun Chen and his team have investigated the origin, transportation and deposition of Asian dust; they used the loess and Red Clay sediments in North China and the aeolian deep-sea sediments in the North Pacific Ocean to reconstruct past environment changes during the late Cenozoic era.

They have focused on three major areas. First, they established the spatial patterns of Nd–Sr isotope ratios, rare Earth element compositions and detrital mineral compositions for the aeolian sediments in the deserts, sand fields, Gobi deserts, loess deposition areas and the Tibetan Plateau. Based on these results, they were able to delineate the source regions of Asian dust and found direct evidence for the close relationship between the isotopic composition of desert deposits and the tectonic settings of the surrounding mountains.

Second, by analysing Nd–Sr isotopes and the geochemistry of detrital monomineral, they found that the direction of short-range transportation of dust in central and eastern Asia is largely determined by the prevailing near-surface wind. Source tracing revealed that the Asian dust has two material sources: the northern margin of the Tibetan Plateau and the Central Asian Orogen. This confirms the importance of mountain processes in the production of silt aeolian particles.

Third, they found that all the sand fields in northern China, an important dust source today, were nearly completely covered by vegetation during the Holocene Optimum, whereas the deserts in northwestern and central northern China were about 5–20 per cent smaller in area at this time. This suggests that the surface conditions of Asian dust source regions changed dramatically during the glacial and interglacial climate variations, and such changes affected regional dust emissions and subsequent feedback to the climate change. Thus, the dust impact should be coupled with the climate system in palaeoclimatic modelling in order to fully understand the forcing mechanism of the past climate changes.

The unique processes of dust deposition and transportation in central and eastern Asia provide a key to understanding the evolution of the Earth's environmental systems during the late Cenozoic. The ambitious research project of Chen and his team focuses on investigating the global dust emission, transportation, deposition and their climatic impact at daily to orbital timescales. They are using high-resolution and high-precision geochemistry and mineralogy techniques to pinpoint dust sources and look for new biogeochemistry proxy indices to better reconstruct past climate and environment changes. Their research will significantly improve our understanding of the behaviour of the Earth's surface environmental systems.

Information sciences

LAMDA, or 'Learning and Mining from Data', is a research team led by Zhi-Hua Zhou, who is an ACM Distinguished Scientist and IEEE Fellow. He is affiliated with the National Key Laboratory for Novel Software Technology of NJU; this laboratory has topped the rankings of the national key laboratories in computer science in China in the past two five-year terms. Zhou's team focuses on research of machine learning and data mining, with the aim of constructing computer programs that can analyse existing data and generalize to future cases.

Zhou and his collaborators developed many innovative techniques that have been applied to various practical tasks including image retrieval, movie annotation, network monitoring, CPU design, medical diagnosis, gene pattern annotation and mobile photograph management. Their innovations have also been incorporated into scientific software such as Matlab and many open-source software systems. Part of their research was awarded second prize in the National Natural Science Awards of China. In that year Zhou also received the Annual IEEE CIS Outstanding Early Career Award, which is awarded to one outstanding researcher under 40 years old.

When Zhou founded LAMDA ten years ago, there were only two faculty members; now the team has grown to 11 faculty and research staff. LAMDA seeks to attract overseas researchers as well as cultivate local talent. Zhou obtained all his degrees at NJU and is an excellent example of home-grown talent.

NANJING UNIVERSITY

Visit: www.nju.edu.cn
22 Hankou Road, Nanjing 210093, China

香 港 大 學
THE UNIVERSITY OF HONG KONG

THE UNIVERSITY OF HONG KONG: INNOVATION, INTERNATIONALIZATION AND IMPACT

The University of Hong Kong (HKU) is a comprehensive, research-focused institution, ranked as one of the top universities in Asia and the world. HKU scholars have made their mark locally, regionally and globally in a wide range of fields. The university supports their vital, cutting-edge research with an English-medium environment that promotes excellence and innovation and a commitment to making a positive impact on society.

"HKU has fantastic research opportunities, but we are not satisfied with leaving it at that," says HKU president, Professor Peter Mathieson. "We need to take the lead on impact because it will make our research better and because we have a moral responsibility to use our knowledge to inform public debate."

A firm foundation for success

HKU attracts some of the best scholars and students from around the world, with nearly two-thirds of its professors and post-graduate students coming from outside Hong Kong. Researchers at the university have the highest citation rate per academic in Hong Kong. Moreover, in 2014, Thomson Reuters ranked 115 HKU scholars — one in ten scholars at HKU — among the top 1 per cent in their fields globally, and named 5 HKU scholars among the most influential in their fields. In addition, 15 HKU researchers have been conferred membership to the Chinese Academy of Sciences and the Chinese Academy of Engineering,

which are among China's most prestigious academic bodies.

HKU scholars have conducted ground-breaking research in fields such as influenza and infectious diseases, geology and geochemistry, economic policy and international finance, drug discovery, energy transmission, the modern history of China, and suicide prevention. In 2014, HKU won the most funding from competitive general research fund grants among universities in Hong Kong for the 13th year in a row. HKU scholars also coordinate the highest number of publicly funded, cross-institutional projects and oversee five Partner State Key Laboratories, which are key components of China's system for science and technology research.

Sustaining excellence

With so much raw talent, HKU is keen to encourage interdisciplinary collaboration. Its strategic research themes initiative provides seed funding in five key interdisciplinary areas: biomedicine, frontier technology, the environment, China and the community. 'Blue-sky research' also remains central to HKU's strategy and receives substantial support.

Regional and international collaboration is another key goal at HKU. The university collaborates with international centres such as the Institut Pasteur in France and the Karolinska Institutet in Sweden, runs joint PhD programmes with King's College and University College in London as well

as the University of Toronto, and has two major research centres in Zhejiang and Shenzhen in mainland China. These research centres capitalize on HKU's critical location at the gateway to China and attract considerable funding.

The university makes a special effort to ensure that the integrity of its output is consistent with the highest international standards. HKU is a leader in Asia in this regard. It provides a compulsory research ethics course for incoming research students and mandatory seminars for staff, and has established a dedicated research integrity fund.

HKU's energetic pursuit of excellence in research has established its standing among the best universities in the world. As the university moves forward, it will focus on enhancing and sustaining that success, as well as broadening its vision.

Three words sum up the future for research at HKU: innovation, internationalization and impact. While much groundwork has been laid for the first two goals, impact is a relatively new target. Research with impact is expected to lead to greater funding opportunities, new collaborations and, most importantly, contributions that will advance society.

香 港 大 學
THE UNIVERSITY OF HONG KONG

www.hku.hk
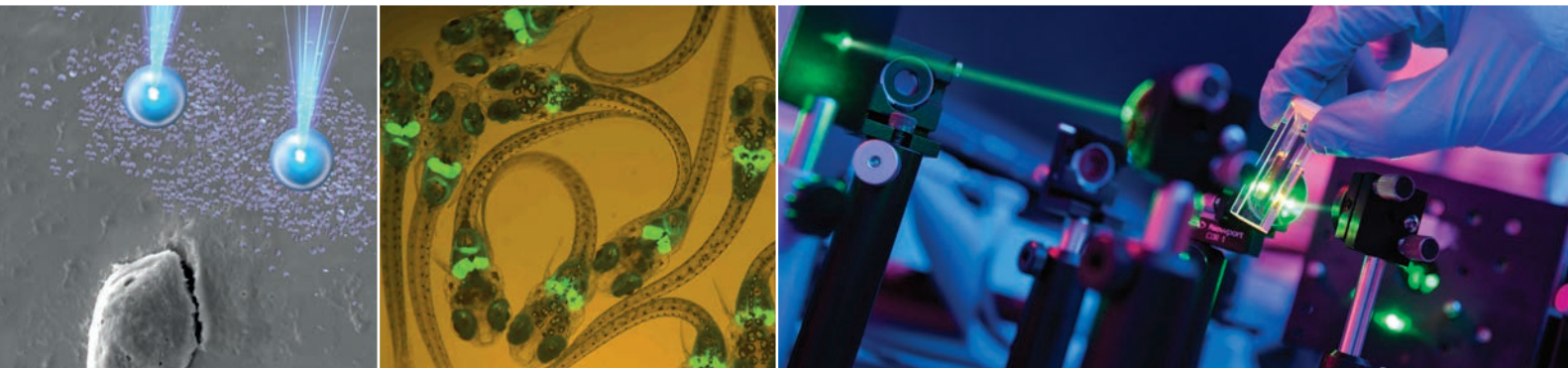
FORESIGHT AND INSIGHT: INNOVATIVE RESEARCH IN EAST ASIA

City University of Hong Kong has established itself as one of the most innovative universities in Asia, pioneering research in fields as diverse as creative media, energy, environment and veterinary medicine, and pursuing wide-ranging projects, from healthcare management to robotics, nanotechnology and superhydrophobic materials.

A relatively young institution — it celebrated its 30th anniversary in 2014 — City University of Hong Kong

Visitors to an installation at City University of Hong Kong can experience the magnificent Buddhist wall paintings inside China's famous Magao Caves in virtual reality.

has risen with remarkable speed in several global rankings. The QS World University Rankings 2014–2015 ranked the University 108th in the world, 11th in Asia and 5th in the world for universities under 50 years of age. According to data compiled by Shanghai Jiao Tong University for the Academic Ranking of World Universities 2014, City University of Hong Kong ranked first in Hong Kong for engineering, technology and computer sciences, as well as for the more specific subject area of mathematics. Globally, the University listed 24th for the broad fields of engineering, technology and computer sciences and 28th for mathematics. The University's College of Business came out second in the Asia-Pacific region in the University of Texas at Dallas Top 100 Business School Research Rankings 2010–2014.

The University's global success has been, in part, a result of its commitment to the four 'i's of integration, innovation, interdisciplinarity and initiative.

Integration

City University of Hong Kong believes that students benefit by learning directly from academics actively involved in research, as well as from teaching staff who immerse themselves in research. This philosophy drives the University's pursuit of the integration of teaching and research.

Innovation

City University of Hong Kong's dedication to innovation in teaching and research is

typified by a recent study tour to one of the world's most dramatic landscapes. Students and faculty at the University transformed data collected during a series of environmental experiments in Antarctica into creative media artwork, which was exhibited in Hong Kong. The result was a perfect merging of art and science in one of the most innovative expressions of fusion research supported by an institution of higher education.

Interdisciplinarity

City University of Hong Kong also believes that the traditional boundaries between disciplines are no longer adequate for advancing research, and that academic institutions need to facilitate further overlap between curriculums of different disciplines. By bringing together experts from across its campus, the University has been able to explore the many facets of interdisciplinarity. One example of this is a recent healthcare management project, which involved experts from the College of Business and the Department of Systems Engineering and Engineering Management. The project so impressed the government of Hong Kong that it was awarded a HK$20.45 million grant from the Research Grants Council.

Initiative

The University has pioneered research into the intersection of art and technology with the creation of the School of Creative Media in 1998, and in 2009 it set up Hong Kong's first School of
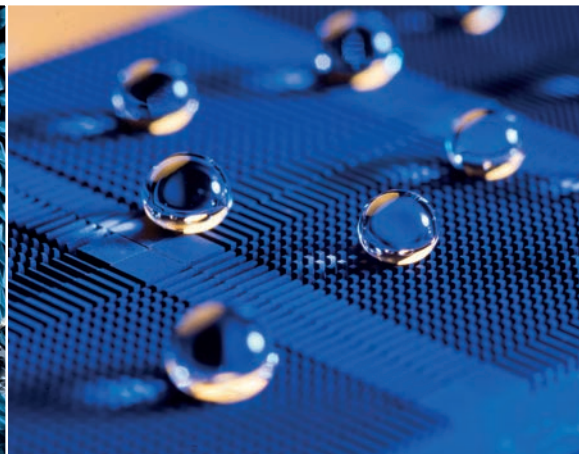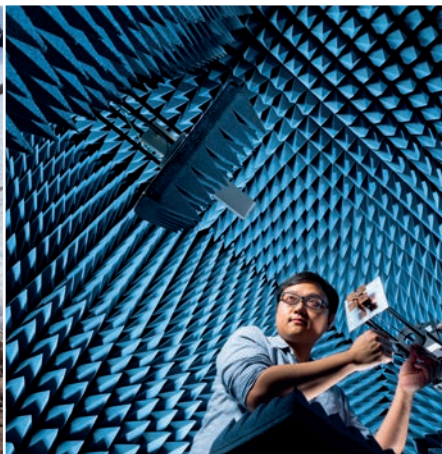
Energy and Environment. Extending this groundbreaking spirit of initiative, in 2014 the University founded Hong Kong's first-ever School of Veterinary Medicine, in collaboration with the College of Veterinary Medicine of Cornell University in the United States. The School of Veterinary Medicine seeks to become a global leader in veterinary education and research, and to establish itself as a centre of excellence for One Health, an approach that recognizes the close links between human, animal and environmental health, ultimately to contribute to improvements in the quality of life in the region.

Government support

Support from the government of Hong Kong and authorities in mainland China is essential for financially sustaining the University and to validate its work. In the Research Assessment Exercise 2014 conducted by the University Grants Committee of Hong Kong, City University of Hong Kong was ranked top among institutions in Hong Kong in terms of the percentage of highly rated research in the areas of building technology, chemical engineering, civil engineering, communications and media studies, materials science, and mathematics and statistics.

Highly cited researchers

Similarly, the talents of individual researchers affiliated with the University have been endorsed and recognized internationally. The University ranked second in Hong Kong in terms of the number of researchers named in the list of Highly Cited Researchers 2014, published by Thomson Reuters. These highly cited scholars work in the fields of chemistry, engineering, physics, mathematics and materials science.

Snapshot: Recent achievements

• **New robotics technology for treating diseases**

An interdisciplinary research team at City University of Hong Kong is collaborating with the University of Hong Kong to develop an innovative cell manipulation and testing system by integrating the technologies of robotics, optical tweezers and microfluidics.

• **Art and technology: Virtual reality applied to ancient Buddhist art**

City University of Hong Kong is pioneering new methods for exhibiting archaeological sites, such as a standout installation that allows visitors to experience in virtual reality the magnificent Buddhist wall paintings inside the famous Mogao Caves at Dunhuang in western China.

• **Food waste into bioenergy**

An innovative process developed at the University uses a mixture of bacteria to generate bioenergy from food waste, heralding possibilities for the development of a viable source of renewable energy.

• **Breakthrough in nanobiology**

A young faculty member in the Department of Biology and Chemistry contributed to the development of a novel nanothermometer for probing the health of living cells. The research, conducted in collaboration with researchers at Harvard University, was published in *Nature* in 2013.

• **Superhydrophobic material**

A team of researchers at City University of Hong Kong has developed the world's most water-repellent surface, which can be used for fabricating anti-icing materials, as well as for self-cleaning, anti-bacterial and anti-corrosion applications. The research was published in *Nature Physics* in 2014.

• **Durable infrastructure**

Joint research between City University of Hong Kong and Massachusetts Institute of Technology in the United States has identified why adhesives holding certain materials together fracture and fail and what could be done to strengthen these bonds, indicating enormous implications for a range of bonded materials.

香 港 城 市 大 學
City University of Hong Kong

Professional · Creative
For The World

City University of Hong Kong
www.cityu.edu.hk

CHINA'S RESEARCH & DEVELOPMENT SPEND

China's investment in research and development (R&D) is second only to the United States. Here, we analyse the data to give a snapshot of how and where the money is spent. **By Xiaole Ni.**

THE BROAD PERSPECTIVE

In 2013, China's R&D spending increased by 15% on 2012 — second only to the United States. Around 75% of spending is by industry, which also contributes the majority of the funds. Looking at the breakdown, the largest category of research is technology development[1]. All figures are in billions of yuan.

Data are from 2013, unless indicated.

2013 conversion rate
1 yuan = US$0.16

1,184.7 BILLION YUAN

CONTRIBUTORS TO R&D

| 883.8 Industry | 250.1 Government | 50.8 Others |

WHO SPENT IT?

| 907.6 Industry | 178.1 Public research institutes | 85.7 Higher-education institutes | 13.3 Others |

SPENDING BREAKDOWN

| 1002.3 Technology development | 126.9 Applied research | 55.5 Basic research |
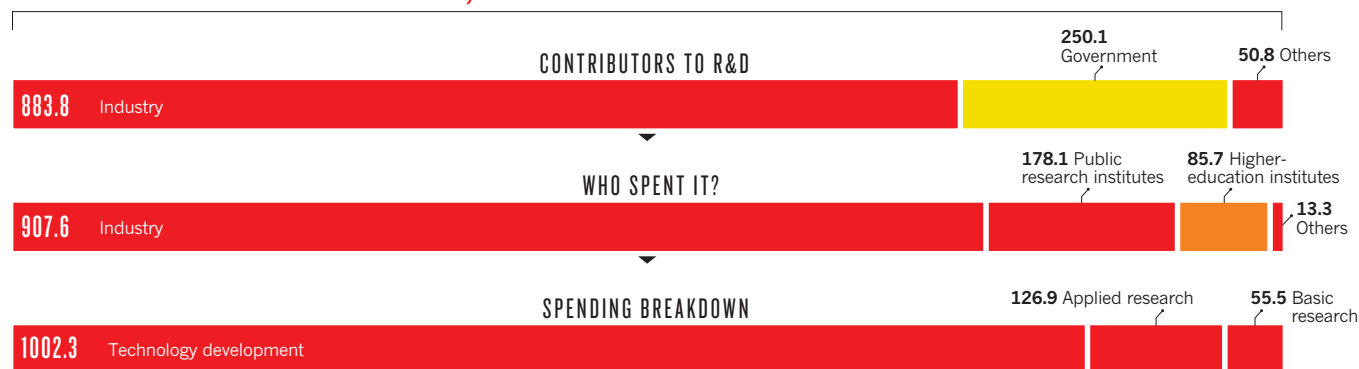
Investment hotspots

There are 31 areas under central government control that together spend mainland China's R&D budget.
For some, this represents a more sizeable proportion of their local gross domestic product (GDP) than for others[1].

Jiangsu, Guangdong, Beijing, Shandong, Zhejiang, Shanghai, Hubei, Liaoning, Tianjin, Sichuan, Henan, Anhui, Shaanxi, Hunan, Fujian, Hebei, Chongqing, Heilongjiang, Shanxi, Jiangxi, Jilin, Inner Mongolia, Guangxi, Yunnan, Gansu, Guizhou, Xinjiang, Ningxia, Hainan, Qinghai, Tibet

R&D expenditure (billion yuan)
150
120
90
60
30
0

R&D expenditure (% of local GDP)
0
1
2
3
4
5
6

Henan's main industries produce a large GDP, but its proportional research spend is low.

Agricultural Hebei spends only a small proportion of its GDP on research.

With its high concentration of research institutes, Beijing's proportional spend is greatest.

Tianjin is positioning itself as an incubator for innovation.

2.08%
Proportion of national GDP spent on R&D

HIGHER-EDUCATION INSTITUTIONS

In 2012, 1,025 universities received a total of 76.9 billion yuan in R&D funding from the government, which they supplemented by raising additional money from other sources. Universities may be under the direct control of government ministries (for example agriculture, education or information) or of local government. Complicating the picture, some reports wrap R&D funding into the larger category of science and technology (S&T) funding, which totals 117 billion yuan and includes operational costs[2]. All figures are in billions of yuan.

45% 29 universities receive nearly half of all S&T funding.

TOTAL R&D FUNDING

76.9

The figure of 76.9 billion yuan is from a different year to the 85.7 billion yuan in 'The broad perspective'.

FUNDS RECEIVED

39.2
26.5
11.2

FUNDS SPENT

30.7 Applied
19.7 Basic
9.1 Technology development

RECIPIENTS

55.1 Project 211
21.8 Rest (581 universities, 333 junior colleges)

China's top 111 universities are part of Project 211, a government initiative to raise research standards in its higher-education institutes. Project 211 includes a subset of 39 universities in the elite Project 985. Project 211 establishments receive the most funding; other universities receive less money and need to raise a greater proportion of funds by themselves.

TOP EARNERS

Tsinghua
Zhejiang
Shanghai Jiaotong
Peking
Fudan
Beihang
Huazhong University of Science & Technology
Tongji
Harbin Institute of Technology

Name of university

S&T funding received (billion yuan) — 0 1 2 3 4

GOVERNMENT FUNDING

There are three main agencies, detailed in shaded pink bars to the centre right, that account for nearly one-third of central government R&D spend[3]. All figures are in billions of yuan.

176.5 Central government spend

250.1 Total central and local government spend on R&D

74 Local government spend

176.5 Given to agencies by central government

107.1 Used by other government agencies | **69.4** Used by the top three funders

Around 70 government agencies fund R&D. Such agencies include the Ministry of Education, the Ministry of Industry and Information Technology, the Ministry of Health, the National Development and Reform Commission, and the Ministry of Defence.

Money spent by the three main agencies also includes additional non-governmental income. Amounts shown for the NSFC represent approved grant funding.

MONEY SPENT ON RESEARCH

Applied 17.8 | Basic 17.7 | R&D 0.6 | Other 5.6

Applied 9.8 | Technology development 8.2 | Basic 4.3 | Other 5.6

Research programmes 15.3 | Talent training 6.3 | Research support 1.9

- Chinese Academy of Sciences (CAS)
- Ministry of Science & Technology (MOST)
- National Natural Science Foundation of China (NSFC)

A complicated picture

There are four types of mechanism in place to allocate funding. The top two types involve around 100 competitive grant schemes that are overseen by 30 different government departments[4].

FUNDING STRUCTURE IN 2015

Competitive, peer-reviewed, investigator-initiated Including the NSFC programme for exploratory research

Competitive, peer-reviewed, mission-orientated Aimed at meeting national S&T challenges

Mission-oriented contract research Including defence-related spending

Operational costs Including for the state key laboratories
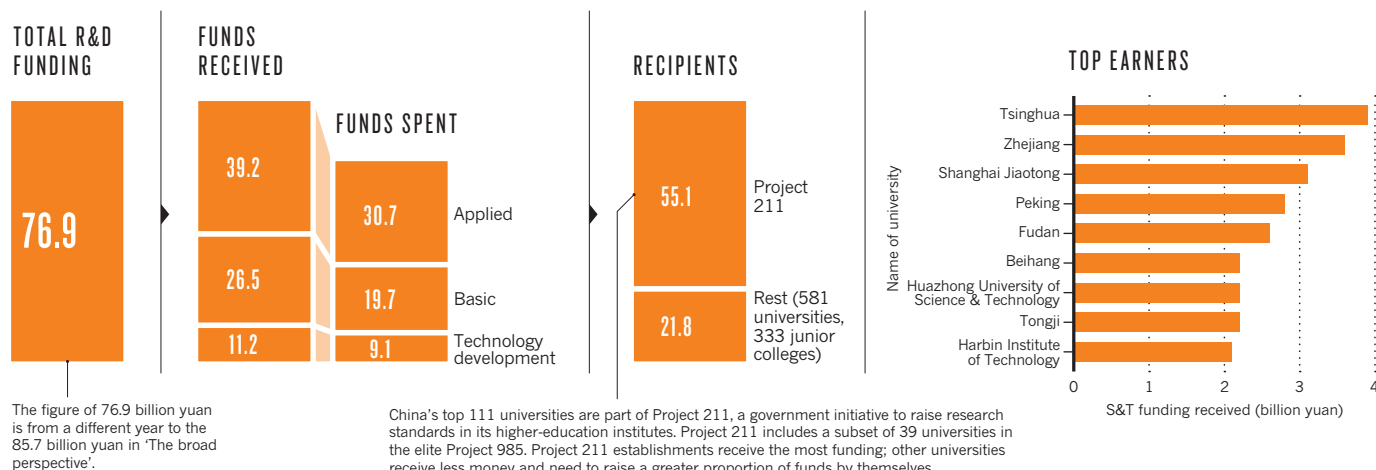
FUNDING STRUCTURE FROM 2017

NSFC programme
S&T major national projects
R&D major national projects
Special innovation fund
Human resources & infrastructure

China's government announced that by 2017, the two competitive grant categories will be modernized and managed under a new science and technology agency into five channels.

Sources: 1. National Bureau of Science and Technology of the Ministry of Finance (2013); 2. Higher Education Science and Technology statistics from the Ministry of Education (2013); 3. China's Ministry of Finance (2013); 4. Sun, Y. & Cao C. Science 345, 1006–1008 (2014).

A panel comprising (L–R) Philip Campbell, Shen Wenqing, David Sweeney, Daniel Hook, Maki Kawai and Li Xiaoxuan discuss China's research evaluation.

SYMPOSIUM OVERVIEW

Raising standards

Growth in the science budget and in research output have been China's key attributes over the past few decades, but the focus is now on how to boost research quality.

BY MICHELLE GRAYSON

Countries the world over are looking to their research base to drive progress in an increasingly competitive environment. But how should research be guided or rewarded to achieve a country's goals — be they academic excellence, as exemplified by breakthrough papers and awards such as the Nobel prize, or measured in terms of economic prosperity and social wellbeing?

"Assessment and evaluation of research is a timely topic globally," said Charlotte Liu, regional managing director for science (Greater China) and education (Asia) at Macmillan Science and Education, the parent company of Nature Publishing Group, when she opened the 2014 International Symposium on Research Assessment and Evaluation in Shanghai, China. "Crucial to this aim is to establish a comprehensive, rational and systematic evaluation framework to guide research investment and

resource evaluation." At the symposium, held in October at the Shanghai Association for Science and Technology (SAST), representatives from academia, industry and government convened to talk about their experiences and to discuss courses of action.

GROWING AND LEADING

In his introduction, Yang Jianrong, vice-chair of SAST, outlined important elements that any research evaluation plan must address. The first, he said, was a "focus on the quality of research benefits, including the need for objective criteria". He stressed that, in a world plagued by resource shortages and environmental degradation, it is important to target innovations that are "beneficial to the planet's long-term stability and sustainable green development".

The pages of *Nature* reflect the substantial growth of Chinese science in recent years, said the journal's editor-in-chief, Philip Campbell. The ultimate aims of research

assessment, he said, were "to incentivize good practice and critical and ambitious thinking". Research has an effect beyond the institution or even the community in which it takes place, with the potential to influence many lives through new technology, new governmental policies and better health care. Therefore, said Campbell, it is crucial to "capture the impact of research in the fullest sense — from both academia and the wider research community".

Any evaluation of research needs to consider the interlocking chain of basic research, applied research, technology development and commercialization, said Zhang Xu, vice-president of the Shanghai branch of the Chinese Academy of Sciences (CAS). "A good system is not only conducive to the development of science and technology," Zhang said, "but it will help our

scientists grow, our education to improve, and help to create benefits for all of mankind and particularly for our environment."

PROMOTING THE UNDERACHIEVERS

Kurt Wüthrich, a biophysicist dividing his time between the Scripps Research Institute in San Diego, California, the Swiss Federal Institute of Technology in Zürich, Switzerland, and the iHuman Institute at ShanghaiTech University, talked about the way in which scientists are evaluated. Wüthrich is best known for sharing the 2002 Nobel Prize in Chemistry for the development of nuclear magnetic resonance spectroscopy. However, during his undergraduate years at the University of Bern he was more successful in his athletic endeavours. He reflected on his track speciality in the conference's plenary address: "For the high jump, one needs to find children with talent and simply measure their achievement. Each attempt gives a clear and final result; good or bad." But such obvious decision-making is not available to science, said Wüthrich. "The result of research, whether it is impact on quality of life or economics, may not appear for years or even decades."

The appointment and promotion of scientists, Wüthrich continued, is often under-emphasized in discussions about research assessment. "If we do not select these scientists well, then we will not get good value out." Wüthrich contends that there are many "talented underachievers" who might need encouragement to make better use of their skills, and who are usually overlooked in favour of people whose main ability lies in maintaining the status quo. These overachieving scientists will not be the ones who produce the breakthroughs, he argued (see page S13).

His message is that, rather than examining researchers' past output, more emphasis should be placed on finding, supporting and retaining the exceptional people. "There are very few who push things forward," he said.

IMPACT IN THE UNITED KINGDOM

David Sweeney provided the counterpoint to Wüthrich's experience — that is, from the perspective of an assessing organization. Sweeney is a director of the Higher Education Funding Council for England (HEFCE), which evaluates research across the United Kingdom. The first task, said Sweeney, is to answer the question: "Why are you assessing?"

For HEFCE, the answer to that question is clear: the evaluations guide the allocation of funding. Most research in the United Kingdom is undertaken at universities, who then choose what projects and researchers to support. "We expect universities to take wise decisions," he explained.

But there is also another crucial question, Sweeney said: "What does research success look like — that is, what are you trying to achieve?" And here, he said, the answer

is very much in flux. The economy of the United Kingdom, as in much of Europe, is struggling. "The government sees our universities and their research as one of the most successful systems in the world and a key part of the approach to returning the economy to balanced growth." The United Kingdom has, he said, "intellectual leadership in the development of new knowledge". But whether this knowledge has a positive impact on society has not been clear. "We assume it has, but do we have the evidence?"

Kurt Wüthrich training in Switzerland in 1956.

For more than two decades, the United Kingdom ran the Research Assessment Exercise (RAE) to evaluate the quality of university research. According to Sweeney, the RAE had a hugely positive effect on the number of publications and citations, and of the quality of this output. "We think we are spending our money wisely," said Sweeney. "We want to spend it even better."

And better, in this context, means more 'impact' — a much-used term at this symposium and in discussions of science evaluation in general (see page S21). Sweeney explained: "Research impact is the demonstrable contribution that research makes to the economy, society, culture, national security, health, public policy or services, quality of life, and to the environment".

For the 2014 assessment, the United Kingdom transformed the RAE into the Research Excellence Framework (REF) and awarded 20% of the evaluation score for an institution on the basis of case studies, which describe the wider (non-academic) impact of research. However, modifying the assessment system is not intended to change the fundamental focus of UK research. "We don't want to discourage curiosity-driven research, but instead to prove that the best impacts come from this type of research," he said.

THE ACADEMY FOR CHANGE

China, unlike the United Kingdom, has no national system for research evaluation. However, the assessment process operating

within CAS — one of the largest research organizations in the world — can be seen as a microcosm for the country, said Li Xiaoxuan, director of CAS's Institute of Policy and Management in Beijing.

China is a newcomer to the international research arena. "It was only 30 years ago that we started to reform," said Li. "We were closed, but we started to open up to a market economy." At the same time, efforts began to modernize industry and science, and the National Natural Science Foundation of China (NSFC) was established in 1986 as the main body for competitive project funding. "But we didn't appreciate how we should guide research through evaluation and assessment," he added.

In 1990, the government began evaluating its national key laboratories, and shortly thereafter CAS began evaluating its own research institutes. The purposes behind the evaluations, said Li, were: to help select the best people through competition; to raise China's science and technology output, in line with international levels; to promote efficient resource allocation; and to bring scientific decision-making into the management of research and development (R&D).

But there were problems with the tools developed for these tasks. "The biggest issue was that there was too much focus on quantity," Li admitted. Scientists would focus on increasing the number of projects they conducted and how many articles they wrote. "It led to very short-term behaviour," said Li, with the knock-on effects of research misconduct and wasted funding.

To provide more helpful incentives, CAS is now moving towards qualitative evaluation (see page S18). Since 2011, CAS has been using the One-Three-Five programme. The name stands for one orientation, three breakthroughs and five major directions. "It is focused on outcomes, not papers," said Li.

By focusing its research institutes on the One-Three-Five plan, CAS is hoping to avoid duplication and create areas of specialization, to help its institutes achieve major breakthroughs and maintain its fast pace of development. Results so far from the 19 institutes evaluated show that around one quarter can be considered world-leading — the highest determination, said Li. One of the core features of the plan is that is uses third-party assessment, including evaluation from international experts. Feedback from such experts should quickly identify any problems in management and provide constructive suggestions. And, as with the United Kingdom's research evaluation scheme, the plan "enables us to keep a balance between basic science and applied research", Li said.

"Assessment and evaluation of research is a timely topic globally."

Wang Minmin, head of chemistry at Eli Lilly's Shanghai-based China research and development centre.

The afternoon of the symposium comprised panel discussions on the broader implications of research evaluation. One of the recurrent themes was how best to support and evaluate young scientists.

"Evaluating young people is different from evaluating older, established scientists," said Li Mengfeng, vice-president of Sun-yat Sen University (SYSU) in Guangdong. "We need to look for potential rather than looking at past achievement."

Shen Wenqing, an academician at CAS, observed that the pervasive conservatism of Chinese society tended to "kill advanced ideas" — ones that have a low probability of success but potentially high impact. Echoing Wüthrich's earlier talk, Shen urged a greater focus on encouraging young scientists to pursue original ideas.

These aims are also of concern to the top institutes in Japan, said Maki Kawai, executive director of RIKEN, Japan's largest dedicated research institution. Young scientists are overburdened by Japan's onerous annual evaluation system, she said. Moreover, the current system does not take diversity into account. "We should not choose just one sort of person," she said, "it is important to have flexibility in the system."

CAS's Zhang raised the issue of how best to evaluate genuine breakthrough ideas. He called on the research community to help develop a new way to determine a paper's quality, one that is not entirely dependent on the journal in which it is published, or on how many times it has been cited. Such a broadening of criteria would be particularly helpful to researchers at the start of their career. "At the very beginning, it is hard to publish in high-impact journals," he said.

Anthony Cheetham, vice-president of the United Kingdom's Royal Society, took up Zhang's point. He noted that "many papers that have recently won Nobel prizes were actually published in second-tier journals". Having papers in these journals only should not be an impediment to promotion, he said (see page S34).

"We need to look for potential rather than looking at past achievement."

Participants started to reach a consensus on this point. Wang Xiao-Jing, associate vice-chancellor for research at New York University Shanghai, added that "these high-impact journals can inform assessment but should not be all of it". When interviewing someone for a position, Wang Xiao-Jing advocates "spending a day reading all their work".

Providing a view from industry was Cory Williams, head of clinical-trial management for Pfizer's Shanghai R&D centre, part of the New York-based pharmaceutical firm's global network. He spoke about the three leading indicators that he looks for in young researchers. The first two concern uniqueness of research and productivity, but the third is perhaps the most important: capacity for collaborative research. "Through collaboration you can reach new skill sets," he said. "These interactions lead to more innovation."

Williams's advice is to "reverse engineer what a distinctive researcher looks like at various points in their career"; then it will be possible to measure and mentor people towards that template from the beginning.

SOCIETAL IMPACT

Another emergent theme was on the challenges in measuring and understanding the impact of all the diverse products of research for society. Campbell gave examples that illustrate how high-impact research is not always published in high-impact journals. "Multidisciplinary teams of natural and social scientists creating solutions for water-stressed cities" is hugely important, despite rarely being published in the leading journals.

Assessment of science is of critical importance to research-funding agencies. According to Chu Junhao a physicist at CAS's Shanghai Institute of Technical Physics, there are four factors that can help funders to determine the societal value of projects: whether the research led to new knowledge; to industrial output; to enhanced technical proficiency; or to development of new expertise.

For some areas of research, the main impact will be in terms of government policy. Lu Yonglong, an environmental scientist at CAS's Research Center for Eco-Environmental Sciences in Beijing, spoke about his speciality. "For us, the point is not to just have papers in top journals, but to ensure that the public understand what the issues are."

A key element in public understanding of science is accessibility of research papers. "We are more likely to have solutions if everyone has access to the relevant information," said Carrie Calder, strategy director for open research at Nature Publishing Group. The value of open access is particularly important when it comes to solving global challenges such as food security, pollution and climate change, which require cross-disciplinary collaboration.

SEEDING NEW COMPANIES

The final headline topic grappled with how to move from research to the creation of new companies. Guo Chongqing, a mechanical engineer at Tongji University in Shanghai, provided a perspective from his four decades at the forefront of engineering design. Guo said that China has come a long way in the past few decades, but that R&D still has its problems. In particular, he said, "there is too much meddling and interference from government".

But there are good signs in certain sectors. "There is a lot of enthusiasm and drive in Chinese universities to contribute to drug discovery," said Wang Minmin, head of chemistry at the Shanghai R&D centre of Eli Lilly, a US pharmaceutical company based in Indianapolis. But key to this field is the ability to learn from failure. "We should emphasize publication and sharing of these stories," Wang Minmin said. At Eli Lilly, she explained, they have "started to celebrate any outcome, positive or negative, to allow people to openly say what they've learned". Such a policy helps to share information and prevent duplication of efforts. Failure, in this context, is also success.

The symposium closed with some reflections on the day's discussions from Chen Kaixian, chair of host organization, SAST. This meeting, involving participants from different scientific evaluation systems, "should help promote understanding of each other, through the blending and collision of ideas", said Chen. As delegates prepared to return home to their companies and laboratories, there was a strong urge to take some of the lessons back with them to have an impact on how science is assessed around the world. ∎

Q&A Kurt Wüthrich
Science regeneration

Nobel laureate Kurt Wüthrich is a Swiss biophysicist who works at the Scripps Research Institute in La Jolla, California, and the Swiss Federal Institute of Technology in Zürich. In 2013, he was appointed as a professor at ShanghaiTech in China, where he is helping to build the institution's science faculty.

What took you to China?

My first trip to the country was in 1983. I was secretary general of the IUPAB [International Union of Pure and Applied Biophysics]. During those years, mainland China was just beginning to emerge from decades of isolation. In the 1970s, there was a historic ping-pong match between US and Chinese athletes to promote diplomacy, and similar things were happening in science.

In the IUPAB and other scientific unions, China was represented by Taiwan; mainland China was isolated. We changed the statutes of the IUPAB so that we could accept two representations for China — one from Taipei and one from Beijing. That way, mainland China could rejoin the international scientific community. This was a very lengthy process and at times frustrating, but eventually it was a highly rewarding affair.

What was the scientific community in China like when you first visited?

There were high points in Chinese science already: at the Shanghai Institute of Biochemistry and Cell Biology, researchers had determined the structure of insulin and had also sequenced a transfer RNA molecule. Those were incredible feats. But the scientists who were in charge in 1983 had barely survived the Cultural Revolution, including physical abuse. They went through terrible times. There were very few who published good work.

How does China's political history affect the country's science research?

China lacks two generations of scientists, who should have been studying during the Cultural Revolution. So although they reopened the universities that had been closed, they didn't have qualified individuals to fill the positions. Even today, there are thousands of poorly qualified professors in universities and colleges. This is, of course, a very difficult situation, and there are some programmes to address the resultant problems. For instance, only a small percentage of professors in Chinese universities are empowered to advise graduate students.

> *"It is very rare to find new Chinese PhDs who want to do a postdoc in China — they all want to go abroad."*

Many new colleges and universities have been founded in China in the past ten years. How are they progressing?

Many have no high-level research faculty. In contrast, I have witnessed the construction of an entirely new campus for ShanghaiTech University, where I am involved in hiring scientists. We try to employ outstanding scientists who are originally from China, most of them are in the United States. We also try to hire non-Chinese scientists. We can offer start-up packages that are much better than those in the United States, yet it is still very difficult to attract top-level scientists. Furthermore, it is very rare to find new Chinese PhDs who want to do a postdoc in China — they all want to go abroad. That's a problem when you are trying to recruit good scientists to build a research team. There is also uncertainty about how long the present 'good life' in China, with all of its support for scientists, will last.

How should Chinese institutions evaluate their faculty members?

What needs to be done is to establish an evaluation process that is based on the quality of science. And this is a difficult thing. Institutions are at a very early stage of introducing proper peer review. Currently they place great importance on publication records. Acceptance of papers in international journals, based on peer review by scientists outside China, is taken as a very positive asset. Because there is no functioning evaluation system within China, the importance given to publication in high-impact journals is enormous. Another problem is that, overall, productivity in China is very high but quality is not. Publication numbers go up, but citation records don't necessarily follow. The solution is strict reviewing of faculty by international experts in science.

China must open its doors to review its system, with help from the international community. Many Chinese faculty want to open up, but a lot of power in China is with the administrators, who often don't write or speak English. Most of the world's top scientific journals are in English.

What can Chinese universities do to train better scientists?

The first recommendation is that all teaching in the natural sciences should be in English. The second is that the most highly talented candidates should be sent abroad, with an eye to recruiting them back as faculty in China.

The third is the most difficult one. That is to take risks and to lavishly support a small number of individuals who promise to do something really new. Scientists who are doing truly original work often don't have a great citation record and don't get published in the leading journals, because it's hard to find someone to evaluate the work. Therefore, the individuals in charge must take time to meet with candidates and have a good sense of what high quality means. Keeping this in mind, science education in China looks forward to a bright future. ∎

INTERVIEW BY MONYA BAKER
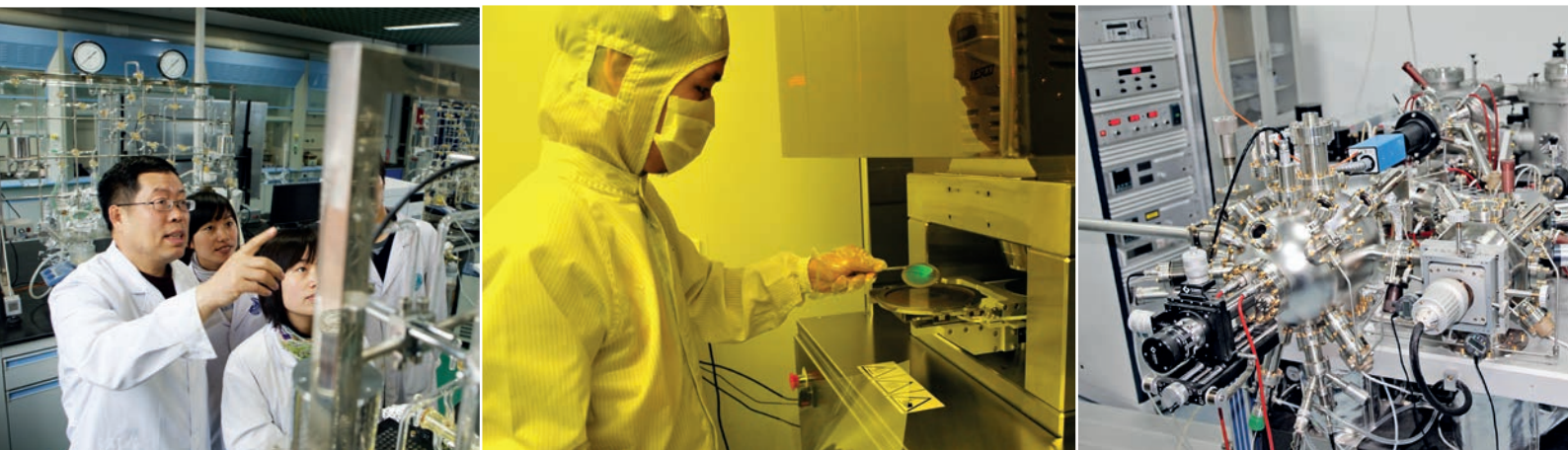
COLLABORATIVE EFFORTS OF CHEMISTRY FOR ENERGY MATERIALS

To achieve breakthroughs in energy science and technology, the Collaborative Innovation Center of Chemistry for Energy Materials (*i*ChEM) was jointly established in September 2012 by Xiamen University (XMU), Fudan University (FDU), the University of Science and Technology of China (USTC), and Dalian Institute of Chemical Physics (DICP) of the Chinese Academy of Sciences (CAS) under the framework of the Collaborative Innovation Plan of the Ministries of Education and Finance of China, the so-called 2011-Plan.

The departments at XMU, FDU and USTC are among the top chemistry departments in China, while DICP is renowned for its leading role in catalysis and energy-related research in China. Such collaboration across multiple disciplines, between different departments and major research laboratories in universities, research institutes and industries is necessary to facilitate world-class research.

*i*ChEM's highly efficient coordination

*i*ChEM proposed and established a multidimensional and innovative "*i*ChEM Efficiency Enhancement System" to encourage flow and collaboration of members. This forms a comprehensive innovative culture and mechanism for remote multidisciplinary collaboration.

The system includes the *i*ChEM Scholars system, the *i*ChEM Sub-Platforms and the *i*ChEM All Member Service system.

The *i*ChEM Scholars system encourages outstanding *i*ChEM talents to interact with researchers of all levels at multiple *i*ChEM research departments. This attracts more talents to *i*ChEM and significantly increases collaboration among multiple units, platforms and mentors — leading to an overall increase of research quality and greater efficiency in human resources.

The *i*ChEM Sub-Platforms make use of the unique academic and physical resources of each collaborative unit to construct distinctive research platforms. The sub-platforms amass complementary resources both from *i*ChEM and externally and allocate them throughout *i*ChEM's multiple units and laboratories. Researchers from different units are able to conduct collaborative research on the sub-platforms to achieve a common research goal.

Attracting talented researchers

As the leader in energy chemistry innovation in China, *i*ChEM holds complete autonomy on talent recruitment, position setup and student admissions quota. *i*ChEM drafts its own recruitment standards, position goals and missions and recruitment protocols. Talented researchers are recruited worldwide and managed via the '*i*ChEM Scholars' system. *i*ChEM provides a world-level research platform, comfortable apartments and

internationally competitive stipends or salary for *i*ChEM Scholars in order to attract expert scientists. *i*ChEM scholars include *i*ChEM Outstanding Professors selected among members of Academies of Sciences and awardees from the Thousand Talents Plan, *i*ChEM professors selected among full and associate professors from renowned universities and research institutes, *i*ChEM research assistants selected from outstanding postdoctoral fellows, and *i*ChEM postdoctoral fellows selected from individuals who have obtained their doctoral degree within the past three years. *i*ChEM also attracts and selects outstanding candidates for doctoral study with excellent mentor teams, challenging research projects, and generous scholarships.

*i*ChEM has recruited a number of well-known experts in chemistry and materials for energy, including ten CAS members, nine Thousand Talents Plan awardees, 39 National Distinguished Young Scientist Grant awardees, and 11 Cheung Kong scholars.

Visit: http://www.2011-ichem.org/en/
Email: 2011-ichem@xmu.edu.cn
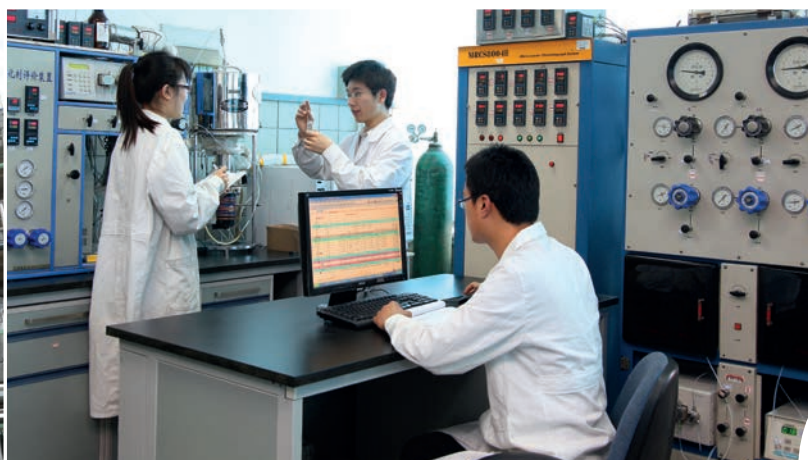
PROMOTING COLLABORATIVE INNOVATION IN CHEMICAL SCIENCE AND ENGINEERING

Established in 2012, the Collaborative Innovation Center of Chemical Science and Engineering (Tianjin) was approved by the Ministry of Education of China in April 2013 to become one of 14 centres in an initiative to promote innovation in research and development and interdisciplinary studies at colleges and universities.

The School of Chemical Engineering and Technology of Tianjin University and the Chemistry Department at Nankai University are highly regarded in China. They are located close to each other and have a long tradition of conducting interdisciplinary research and education. The centre is founded on collaboration between the School of Chemical Engineering and Technology at Tianjin University, the College of Chemistry at Nankai University, the Institute of Process Engineering of the Chinese Academy of Sciences, the Sinopec Group (a Fortune Global 500 company) and the Tianjin Bohai Chemical Industry Group. The centre conducts research in the areas of design and green synthesis of functional materials and highly efficient, clean conversion and utilization of energy and resources. It performs both fundamental and applied research to solve key issues in the fields of energy, resources, environment and healthcare.

The Collaborative Innovation Center of Chemical Science and Engineering (Tianjin) has gathered many world-leading scientists in the fields of chemistry and chemical engineering. In the area of design and green synthesis of functional materials, these scientists mainly focus on studying structural effects and modulating the properties of functional molecules as well as developing new, highly efficient green reactions, reagents and catalytic systems. Researchers at the centre also investigate the catalytic modes, chemical reaction rules, the efficiency and selectivity of the creation and/or conversion of chiral materials and the precise and perfect synthesis of chiral drugs and materials. Furthermore, they study the design, synthesis and application of artificial biosystems, which are associated with synthetic biology, genomic and pesticide chemical biology and organic synthetic chemistry of bioactive molecules.

In the areas of conversion and utilization of energy and resources, the centre mainly focuses on studying catalytic theory and engineering technology related to highly efficient and clean conversion of coal, synthesis gas and carbon dioxide. For the development of new energy and resources, it focuses on studying design theory and engineering technology of high-efficient photoelectric conversion materials and devices, and highly efficient light-catalysed reactions.

Visit: www.tju.edu.cn/2011/english/

Information for postdoctoral applications for the Collaborative Innovation Center of Chemical Science and Engineering (Tianjin)

The Collaborative Innovation Center of Chemical Science and Engineering (Tianjin) invites postdoctoral candidates from around the world to join its team.

The centre will provide a good working environment, research and development facilities, as well as professors for supervising projects. The total annual salary will be up to 250,000 RMB. An extra bonus will be awarded to exceptional candidates, who will be recommended for appointment as a staff member at Tianjin University or Nankai University.

Candidates should possess a doctoral degree from a well-known university or institution (either in China or overseas), good physical and psychological health, and an excellent scientific research track record. They should be a team player with an innovative spirit.

Candidates should have a research background in functional molecule synthesis, synthetic biology, catalytic science, highly efficient and clean conversion of coal, synthesis gas and carbon dioxide, clean energy, synthesis and application of chiral molecules, high-efficiency photoelectric conversion materials and devices, high-energy fuel, or other areas in chemistry and chemical engineering. Outstanding candidates will receive competitive advantages.

When applying, please provide a resume (which should indicate the level of scientific research conducted, present and potential ability, and a self-assessment), a proposal for research to be conducted at the centre, two or three letters of recommendation (including one from a doctoral or postdoctoral supervisor), a document outlining representative research achievements and scanned copies of proof of identity and degree certificate.

Please send an email with the subject "Graduate school [name of major-research direction]" with the above materials attached to hxhg@tju.edu.cn.

For more information, please visit http://hxhgxt.tju.edu.cn. We look forward to hearing from you.

DALIAN INSTITUTE OF CHEMICAL PHYSICS, CAS
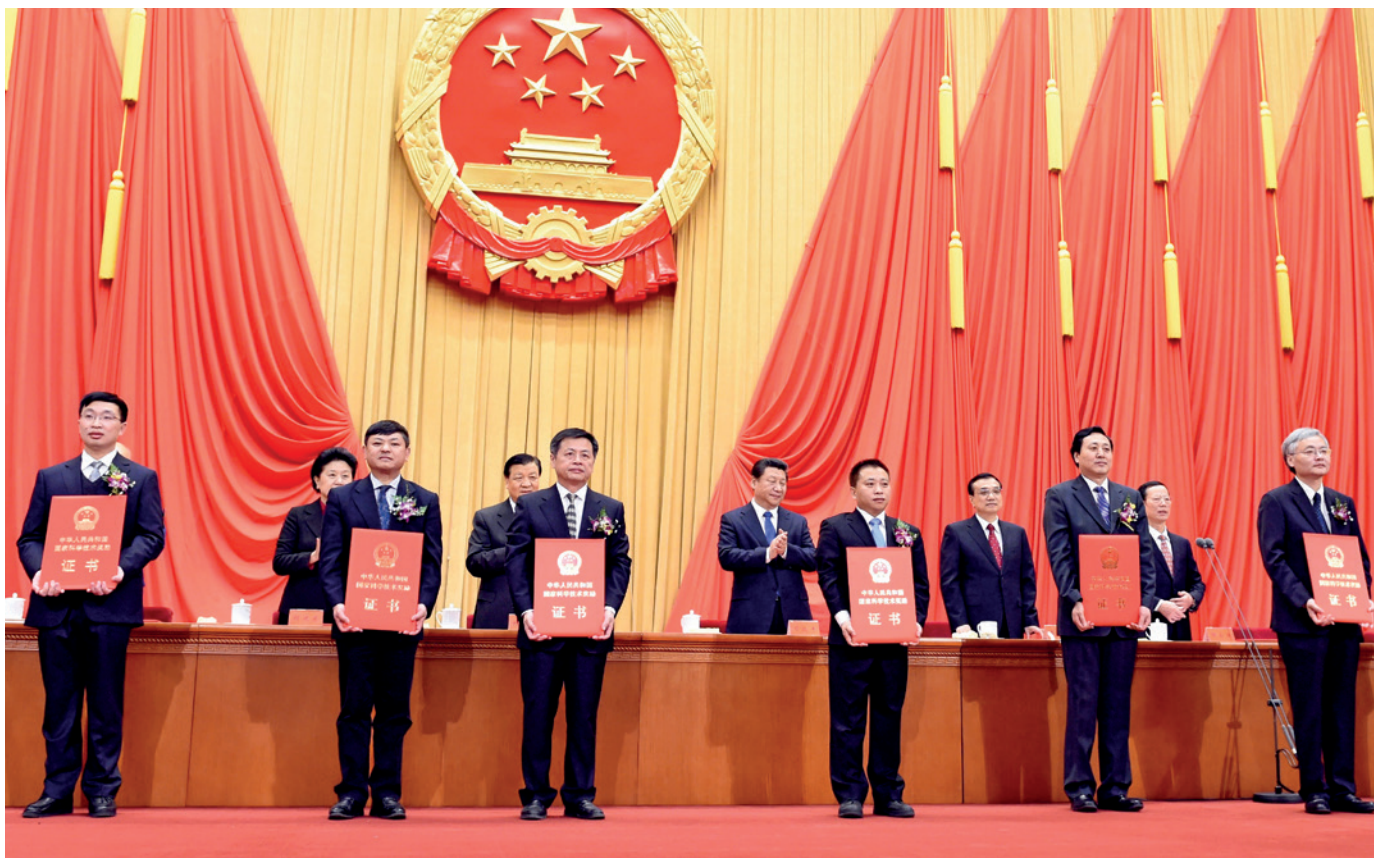
Liu Zhongmin (front row, third from right) receives the prestigious First Class State Technological Invention Award from President Xi Jinping in January 2015.

EVALUATION

Moving away from metrics

The Chinese Academy of Sciences is changing the way it assesses its research. But finding the right balance in such a big organization is a tough job.

BY HUANG KUN

"I never thought it would be awarded by President Xi," says Liu Zhongmin, a chemist at the Chinese Academy of Sciences (CAS), a country-wide network of institutions with headquarters in Beijing. "It's a glorious moment. It is the symbol of social recognition of our research, on which my colleagues and I spent more than 30 years."

Liu is referring to the First Class State Technological Invention Award that he received from China's President Xi Jinping at China's annual National Science and Technology Awards Conference on 9 January 2015. His award-winning invention is dimethyl ether/methanol to olefins (DMTO) technology, which allows the chemical industry to substitute coal for oil in many applications. That is profoundly important for China, which is poor in oil but rich in coal. Liu and his CAS

colleagues invented the process and pioneered its use in industry. The total production capacity for olefins from 7 DMTO plants in China reached 4 million tonnes per year in 2014.

Encouraging researchers such as Liu to make breakthroughs like DMTO technology is the goal of CAS's new assessment system. "That is a good model for constructing the major outcome-oriented system for evaluation at CAS," says Li Xiaoxuan, director of the Management Innovation and Evaluation Research Center at CAS. The centre, based in Beijing, is the academy's official think tank for research assessment.

And CAS is big. According to its latest report, CAS has 56,000 researchers and technicians, plus another 12,000 non-technical staff. It consists of 104 research institutes across China, as well as universities, companies and other entities. In 2013, through a combination of government

funding and other sources, CAS had budget of 41.9 billion yuan (US$6.8 billion), most of which was spent on research and development (R&D) (see page S8).

In the past, CAS has evaluated its research output with quantitative metrics and rankings. But now it is changing the way it operates to give a more nuanced appreciation of progress.

GIVING IT BACKBONE

The core of the current evaluation system at CAS is known as the One-Three-Five plan. "It's difficult to have a panorama of everything in this big organization," Li says. "But now many things can be connected to this backbone."

One-Three-Five is a guide to the way that each CAS institute should operate. That is, each institute should stake out for itself one major research area where it positions its core competencies, set out to make three major breakthroughs in the next five to ten years and

map out five specific research priorities that will provide competitive advantage.

Reforms that led to the One-Three-Five plan began in 2011, when chemist Bai Chunli started his term as president of CAS. He told *Nature* at the time that his plan was to "reduce the frequency of research evaluation while improving its quality" (see go.nature.com/gajb1x). Moreover, he said, research would be judged "based on national needs and socio-economic benefits".

The following year, Bai expanded on his ideas in an article in the *Bulletin of the Chinese Academy of Sciences* (see go.nature.com/xgnkde). In the article, Bai argued that as the growth of China's economy slowed from the frantic pace of the past two decades, it was crucial to also shift the focus of research towards innovation. Thus, the evaluation system at CAS should encourage researchers to solve scientific problems, create new research fields and make major technological advances.

FROM QUANTITY TO QUALITY

One of the main differences between the current and past evaluation systems at CAS is the extent to which quantitative measurements are used. For much of the past decade, institutes were ranked by combining a number of indicators, each with a different weighting, in various complicated formulae. There were 24 basic indicators, including the number of publications and citations, amount of funds raised, awards gained and patents issued.

This heavily quantitative assessment approach led to what Li calls the Science Citation Index (SCI) phenomenon — an abnormal emphasis on increasing the number of papers published in journals in the SCI, even if the impact of such research on industry and society was low. Li believes that the new evaluation system will be more effective at aligning research incentives with the country's overall needs. "The shift of emphasis to qualitative assessment will encourage major outcomes that really matter," he says.

The quantitative system had winners and losers. "Our institute always ranked at the top," recalls Liu proudly. He joined CAS's Dalian Institute of Chemical Physics (DICP) in Liaoning province as a postgraduate in the 1980s. Three decades later, he is now the deputy director of DICP.

Liu believes that DICP is still one of the top five performing institutes at CAS, despite the lack of quantitative ranking. Although the basic indicators are still collected, the numbers are no longer used to calculate a rank order. Instead, each institute submits its data to CAS headquarters, which sends back a report. "That's very useful for us to analyse variations and so adjust our work," Liu says. These indicators now form the foundation of the One-Three-Five plan, and assessment is in the form of two sections. The first section is an expert assessment — a form of peer review

A dimethyl ether/methanol to olefins complex in the Shenhua Baotou plant in Inner Mongolia.

— to check on progress towards the strategic goal, which takes place once every five years.

The DICP's expert assessment was conducted in 2013. The experts recognized the DMTO technology as one of the institute's major breakthroughs, which qualified the project for additional funding from CAS. "And we have many research areas that were judged to be in the first tier in the world," Liu adds. First tier being the equivalent of 'world-leading' research; there are four tiers for rating research from an international perspective.

From 2012 through 2013, CAS conducted international expert assessment of 16 research institutes, including DICP. Experts invited ranged from journal editors to chairs of international societies and associations. They assessed 90 research areas within the 16 institutes, judging 26 to be in the first tier and 53 to be in the second.

INSIDE INSTITUTES

To supplement these overarching evaluations, each CAS institute performs its own assessment of its research groups and individuals. The institutes aim to stimulate competition and ensure that the best researchers are retained.

DICP is one of the biggest institutes at CAS and employs more than 1,000 staff in around 80 research groups. DICP has been performing its own internal evaluation system for many years. In the past, DICP's indicators included quantitative elements, but now, just as CAS is reforming to be more qualitative, DICP is also moving in that direction.

"We have cut some indicators and now are using four basic criteria: the clearness of research direction; the closeness to our One-Three-Five strategy; the progress of their work; and the potential for future development," Liu explained. "Qualitative judgement based on these criteria comes from an academic committee — in essence, a peer review." And these internal assessments can have high stakes.

DICP, for example, evaluates and ranks all of its research groups every two years. "The bottom 5% are eliminated," says Liu.

Not all CAS institutes are moving as quickly or decisively as DICP. In north Beijing, near the stadium that hosted the 2008 Olympic Games, there is a cluster of 12 CAS institutes. *Nature* talked to a scientist who works in one of the institutes, but who wishes to remain anonymous for fear of repercussions for speaking out. He is an associate researcher in his thirties and is one of an estimated 14,000 'middle-class' professionals at CAS. He says that despite these recent changes at the top level of CAS, his evaluation — and that of the many others like him — still relies heavily on the old system of indicators.

"The shift of emphasis to qualitative assessment will encourage outcomes that really matter."

"I have received money according to the number of papers I published, with additional money for each citation of the paper," he says. This kind of quantitative reward system is still very much in existence in many institutes, he adds, they vary only in the details. "Papers bring us monetary rewards, which are given at the end of each year and are typically equivalent to three or four months' salary," he says. "I have known some highly cited papers that brought the authors rewards worth more than a whole year's salary — or even of several years' salaries."

These indicator numbers are also important references for career advancement within CAS. Papers, citations and funds received are considered 'hard assets' to show to the academic committees that evaluate individuals for promotion. Chinese society values relationships — knowing the right people is important and the academic circle is no exception. Therefore, relying on hard assets is one way to make sure that evaluations are not unduly influenced by
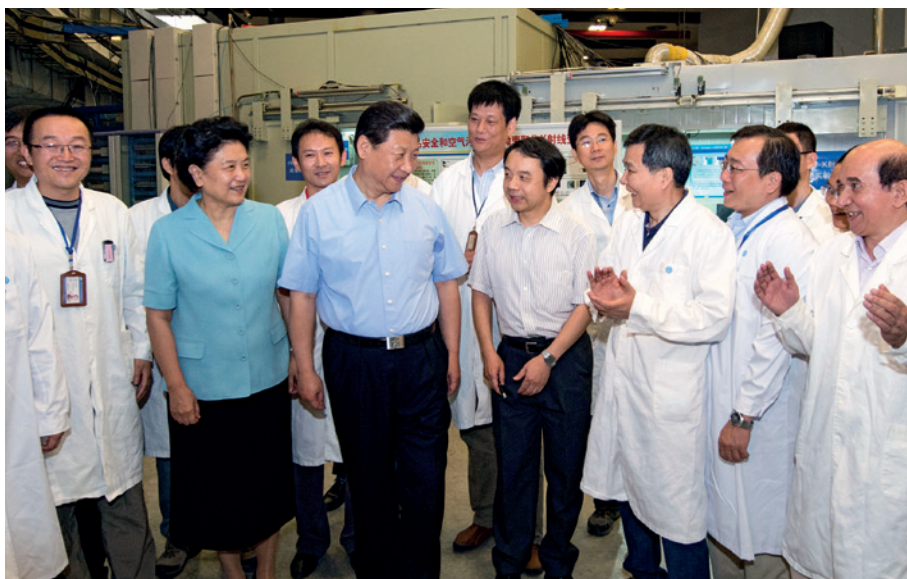
China's President Xi Jinping (third from left) visits the Chinese Academy of Sciences in Beijing in 2013.

scientists' friendship circles. So there is certain rationality in utilizing these numbers, leaving the balance between quantitative and qualitative methods to be a delicate issue.

"Inviting foreign experts to participate in the assessment section of the One-Three-Five plan is a good attempt to solve this issue," says Mao Shude, an astrophysicist at the National Astronomical Observatories of the Chinese Academy of Sciences (NAOC), based in Beijing, who spent 22 years studying and working in the United States and Europe. Mao explains that a group of high-profile international experts will provide a more objective evaluation as they have fewer conflicts of interest or established relationships than domestic ones.

ROOM FOR IMPROVEMENT

Mao, who recently took a joint position at Tsinghua University in Beijing, says that the reform of the evaluation system at CAS is "on the right course". But compared with similar processes at international institutions, he says, there is still room for improvement.

"At foreign universities, the contribution of a person consists of three parts," only one of which is their research, says Mao. The second part is teaching, which has been largely absent at CAS but is becoming more important since the University of the Chinese Academy of Sciences started recruiting undergraduates in 2014. The third part, Mao says, is scientific cooperation and the research environment. "Groups within CAS institutes tend to operate separately, which is not conducive for an interactive and cohesive atmosphere."

Mao returned to China in 2010 as part of the national '1,000 Talent Plan', which recruits high-level experts — both Chinese nationals and foreign professors — from overseas and offers them well-funded academic positions to carry out their research in China. But, he

says, opportunities at Chinese institutions should be based on academic merit, not just experience gained overseas. "Although I am a beneficiary of this system, it's not fair for some domestic researchers who did similar or even higher-quality work but were not recognized as such." This situation is changing, however, as the Chinese government has just started a similar talent programme (the '10,000 Talent Plan') for domestic researchers, to convince them to stay in China rather than relocating overseas, notes Mao.

Another area for improvement is the way in which the evaluation considers the specific needs of female researchers, says Mao. He notes that the percentage of females in permanent positions in his division, from assistant researcher to the highest levels, is just below 15%. "I have two daughters and I am very sensitive about gender equality," he says. "It should be taken into account that women often bear additional family responsibilities."

In the Gravitational Lensing and Galaxy group at NAOC, where there are more than a dozen researchers, Wang Yuting is the only woman. Although she feels she is treated as equal to her male colleagues, she is aware that most female postdoctorates like her are at an age at which they feel they have to choose between research and family life. "Some female scientists did give up research because of family and kids."

According to the United Nations Educational, Scientific and Cultural Organization, 30% of the world's science researchers are women. Although specific data are not available, the situation is broadly similar in

"Groups within CAS institutes tend to operate separately, which is not conducive for an interactive atmosphere."

China and is more acute in the more senior positions. In 2011, the government set a target of increasing the percentage of women in all high-level professional positions to 35%, including a special section devoted to improving the number of women as career scientists. For this to be achievable, the evaluation system would need to be more flexible, as it is in many Western countries, to allow for a career break and support when women return after childbirth, says Mao.

MORE LEADS

With regards to determining the trajectory of CAS, the most important voice is that of China's President Xi. The president visited CAS in July 2013 and in a speech suggested that CAS should aim for "four leads": the development of technology; nurturing the best minds; contributing to science and technology policy; and becoming a first-class international research organization.

In 2014, based on Xi's instructions, CAS started a project called Lead Action. It proposed that its institutes be divided into four categories, each with different evaluation systems based on their missions. These are: institutes that aim to meet the needs of national strategy and industrial development should be evaluated by the people and organizations who use the technologies; institutes that focus on academic achievements should be evaluated by an international academic peer group; institutes with big scientific facilities should be evaluated by users and colleagues; and other institutes with subject specialities should be evaluated by experts in the same field.

Lead Action is gathering momentum. In February 2015, CAS released its latest guidelines for the Academy (the last revision was in 2002). Alongside the four leads are three orientations to ensure that CAS is matching its output with global science and technology efforts, the major needs of the country; and the "main battlefields of the national economy".

The outlook of the evaluation system at CAS is evolving. There are still issues to address in the way that different groups are treated, in ensuring that the various institutes match the overall strategy, and in finding the balance of qualitative and quantitative measurements. But the direction is clear. As Liu explains, flaws in the former evaluation system gave unclear signals to scientists as to what was the most valuable outcome. "There were multiple batons that pointed in different directions," he says. Now, he believes that the modified assessment system has brightly illuminated the goal: "Chinese scientists should do things that are useful to China first of all, then they could consider the rest of the world." ∎

Huang Kun *is a science correspondent based in Beijing.*

Q&A David Sweeney

The numbers game

David Sweeney, a director of the Higher Education Funding Council for England (HEFCE), outlines the importance of accurately assessing the benefits of academic research and the dividends it can bring.

How do China and the United Kingdom differ in their methods of evaluating and funding research?

For half of the higher-education assessment and funding in the United Kingdom, we don't make judgements about individual projects. We fund as a block grant without specifying what the money should be spent on. This means that a university can earn money because of its excellence in physics and spend it on drama. My understanding is that the Chinese approach involves more direction concerning the disciplines and institutions that the money will be spent on.

The United Kingdom's competitive research grant system is, of course, based on judgements, but those judgements are made solely by discipline experts. The only role for centralized planning by government is in the area of capital expenditure. Overall, the government and research funders cede the judgements to either discipline experts in the research councils or to university management.

Should China take inspiration from the United Kingdom in its efforts to reform its research assessment and funding systems, or vice versa?

The UK system is relatively mature. We've been conducting research in universities for a very long time, and we've been doing it in the way we do it now for at least 25 years. We are not on the whole trying to build research capacity.

Our system serves us well and suits our culture, but that doesn't necessarily mean it is right for another country. The real question is: should you do things differently in a research system that is in the process of establishing itself? At the 2014 International Symposium on Research Assessment and Evaluation, held in Shanghai and co-hosted by *Nature* (see page S8), I was very taken by some of the honest comments about how twentieth-century Chinese history affects the nation's research structures, notably how the age cohorts of the researchers in China essentially reflect historical events such as the Cultural Revolution. We have nothing like that in our recent history that has changed the way we do things in the United Kingdom. It's quite tricky to compare assessment and funding systems when the environments are so different.

What other countries have unusual ways of directing and funding their research efforts?

One extreme example is Singapore, which has a highly planned and directed system. Singapore is seeking to build very strong research organizations and is offering significant incentives to encourage the best brains to go and work there. Another interesting case is

Australia, which targets its research resources in its strongest areas, such as environmental science, or those in which there is considerable national need, which in the recent past has included production methods that can accommodate competing demands for soil and water.

You also have to make a judgement about whether one of the purposes of research is to support broader innovation and business. The United Kingdom is second only to Switzerland in the Global Innovation Index 2014 (an annual ranking of countries by innovation metrics). So depending on what they are trying to achieve, countries may need to look beyond their research funding systems, perhaps to associated ways to translate research into jobs and other economic benefits to achieve the biggest impact.

Should other countries be worried that China has increased its research and development funding by an average of 23% per year for the past decade?

The United Kingdom is fortunate enough to be extremely productive, which is partly because of the maturity of our system. If we want to maintain our current position, then we have to ask: 'Are we are investing enough?'; and 'If we maintain our current level of investment, will the best people be attracted to work elsewhere because of the greater investment that others are making?' Our success is partly due to the people we attract from other countries, so yes, there is a threat posed by others having more funds for investment. We have to consider at every stage whether we have sufficient resources to attract and support the best people.

To what extent is China's rapid rise in high-quality research output a direct consequence of its sustained increases in funding?

I'm sure that increased investment supports that and I'm also sure that there are other factors in play, but we don't yet have enough evidence to understand what those factors are. It would be helpful to understand not just the role of increased investment on results, but also how you get those results. We could all learn about using money more effectively.

Should other countries follow China's example by increasing research funding?

You've got to do analysis on what the outputs of your research are, whether that is publications or new technologies. It's always good to invest in developing new knowledge but there is competition for scarce funds from other areas such as education, health and defence. You can only determine whether the potential benefits are worth a proposed increase in investment based on the particular problems a country is facing and on the productivity of your research system. If a country's productivity is low, it might be difficult to justify additional investment in academic research. ∎

INTERVIEW BY NIC FLEMING

DALIAN INSTITUTE OF CHEMICAL PHYSICS: A PROVEN RESEARCH CENTRE OF EXCELLENCE

Dalian Institute of Chemical Physics (DICP) is located in the beautiful port city of Dalian, China. Since its founding in 1949, DICP has built up an impressive portfolio of achievements in the basic and applied sciences that have directly impacted the economic and technological development of China. Traditional areas of excellence at DICP include catalysis, chemical engineering, chemical lasers, molecular reaction dynamics, organic synthesis, modern chromatographic techniques and biotechnology.

DICP is composed of eight laboratories, each having unique research approaches and imperatives. While having distinct and diverse capabilities, technologies and equipment, all of the laboratories work together in unison to accomplish important national directives.

With over 1,000 staff members, 950 graduate students and 100 postdoctoral scholars, the research efforts at DICP are extensive. Many outstanding scientists and engineers have chosen the institute as a home base where they can develop strong skills and enhance their domestic and international reputations. DICP currently boasts 17 members of the Chinese Academy of Sciences (CAS) and Chinese Academy of Engineering (CAE) and 3 members of the Academy of Sciences for the Developing World.

Making an impact
Each year, programmes at DICP are becoming increasingly recognized globally, and domestic and international collaborations are expanding. The quality of the research output at DICP is reflected in part by the number of research articles published in high-level journals. In 2014, DICP researchers published eight articles in the prestigious scientific journals *Science* and *Nature*.

DICP is developing an impressive number of valuable technologies for industry. In the past 15 years, DICP has been granted over 2,000 patents, placing DICP ahead of all CAS institutes in terms of number of patents awarded. DICP's innovative research focused on enhancing China's national economy and security has garnered widespread praise, with DICP faculty being awarded over 85 national awards and more than 230 provincial-level awards, including the National Supreme Science and Technology Award, which was awarded to Professor Cunhao Zhang in 2014.

Focusing on modern energy challenges
DICP is strongly focused on addressing challenges related to sustainable energy through cutting-edge, collaborative research in all of its laboratories. With a tradition of excellence in basic and applied research and interdisciplinary innovation, DICP is poised to continue to make important advances in a number of research fields. It is placing special emphasis on energy-related research, including the development of cutting-edge theories and technologies for optimal fossil energy use, high-efficiency chemical energy conversion and renewable energy to meet national strategic and global needs.

Dalian National Laboratory
The largest laboratory at DICP is the Dalian National Laboratory for Clean Energy (DNL), which is China's first national laboratory dedicated to energy research. DNL, which consists of ten divisions covering research areas such as fossil energy conversion to hydrogen energy and advanced materials, is already playing a prominent role in China's national economy and security.

Clean coal technologies
Because coal is China's dominant energy source, the clean use of coal is critical for the nation's sustainable development. Researchers at DNL are contributing to the country's clean coal effort through the development and refinement of a key process called DICP methanol to olefin (DMTO). This process efficiently converts methanol obtained from coal syngas (synthesis gas) to small alkenes, primarily ethylene and propylene. It has been commercialized through strategic collaborations between DICP and major industrial partners, including SYN Energy Technology Company Ltd and Luoyang Engineering Corporation, Sinopec. Several plants around China are producing around 4 megatons per year of olefin products. In 2014, the DMTO process

was awarded first prize in the prestigious Chinese National Technological Invention Awards. DICP has also made important progress on a variety of other conversion technologies, including DICP methanol to propene (DMTP), syngas to ethanol, syngas to high alcohols and synthetic natural gas. As with the DMTO process, DICP will soon be industrializing these technologies.

Alternative energy and energy storage

In recognition that non-fossil fuel energy sources will be necessary to meet China's future energy demands, DNL researchers are also making pioneering advances in alternative energy technologies, including solar and biomass energy, fuel cells and energy storage. Scalable energy storage solutions for commercialization are a research priority at DICP. Major innovations (with over 100 patents filed to date) in flow battery materials, structure design of stacks and systems, and integration technologies have already been made. In 2013, DICP and an industrial partner demonstrated the successful operation of the world's largest 5 MW/10 MWh all-vanadium flow battery storage system for a 50 MW wind farm in China.

State Key Laboratory of Catalysis

Advances in catalysis are unquestionably one of the most critical components of energy research. Using state-of-the-art experimental and theoretical techniques, the DICP State Key Laboratory of Catalysis is investigating catalytic reactions, creating advanced catalytic materials and developing *in situ* characterization and monitoring techniques that are impacting energy research, as well as environmental and synthetic chemistry. Detailed probing of catalytically active phases, active sites and reaction mechanisms has led to several important insights and discoveries in nanocatalysis, including the synthesis of single-atom catalysts and the discovery of the interface confinement effect.

State Key Laboratory of Molecular Reaction Dynamics

Major breakthroughs in energy research and many other research areas, such as environmental and biological chemistry, can be achieved through detailed understanding of the most fundamental chemical processes. The State Key Laboratory of Molecular Reaction Dynamics (SKLMRD) is one of a handful of facilities in the world possessing both advanced experimental equipment and sophisticated computational capabilities. Using crossed molecular beam techniques, researchers at the SKLMRD combine experimental results with complimentary accurate quantum dynamics calculations to gain deep knowledge about important chemical reactions in the gas phase and at the gas–surface interface at the atomic, molecular and quantum-state-resolved levels. Knowledge of such mechanistic details can lead to new ways to approach technology development, such as designing catalysts for hydrogen production.

CAS Key Laboratory of Separation Science for Analytical Chemistry

DICP's expertise extends beyond energy to a wide range of research fields that address current global research challenges. The CAS Key Laboratory of Separation Science for Analytical Chemistry is a leader in the development of new methods for sample preparation, separation and identification of species in complex samples, with a particular strength in the development of chromatography-based techniques and theory. The laboratory has also achieved recognition for developing and refining techniques and instruments for monitoring environmental pollutants; these include advanced air sampling technologies and detection of trace pollutants.
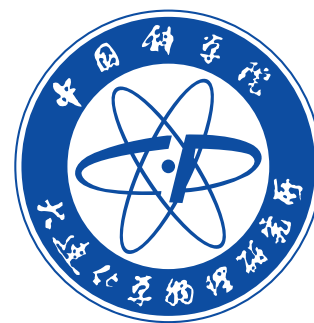
Dalian Institute of Chemical Physics
Chinese Academy of Sciences
457 Zhongshan Road
Dalian, China 116023
Tel: +86-411-84379175
E-mail: info@dicp.ac.cn

www.english.dicp.cas.cn

SHANGHAI CENTER FOR SUPERCONDUCTIVITY CHINESE ACADEMY OF SCIENCES

The Shanghai Center for SuperConductivity (SC^2) was established in 2012 as a dedicated center of the Chinese Academy of Sciences (CAS), associated with the Shanghai Institute of Microsystem and Information Technology (SIMIT). It hosts approximately 75 staff members, including 29 full and associate professors. The centre performs excellent multidisciplinary research on superconducting electronics and seeks to bridge the gap between fundamental research and real-world applications. Its research focus includes superconducting devices and circuits and system integration as well as applications of superconducting devices, novel quantum materials, electronic structure characterization and first-principles calculations.

SC^2 operates a dedicated superconducting device nanofabrication facility equipped with advanced electron-beam lithography and an i-line stepper. It is also constructing a comprehensive research platform for the analysis of electron structure that will employ two state-of-the-art synchrotron beamlines.

Major research directions

Superconducting quantum devices: Conducting research into device physics and developing superconducting quantum interference devices (SQUIDs), superconducting nanowire single-photon detectors (SNSPDs), novel quantum devices and circuits.

Frontier applications of superconducting devices: Exploring applications of SNSPDs and SQUIDs, such as quantum information, communication and imaging based on advanced photon-counting technology, biomagnetic imaging, ultralow-field magnetic resonance imaging and geophysical prospecting.

Quantum materials: Investigating novel quantum materials including graphene, topological materials and superconducting hybrid structures.

Electronic structure characterization of novel materials: Conducting research involving *in situ* soft X-ray surface characterization, analysis of the electronic structures of epitaxial materials, and first-principles calculations.

Talent recruitment

Position 1: Superconducting devices and circuits

Job description: Develop superconducting devices and circuits including single flux quantum (SFQ) devices, transition edge sensors (TESs) and qubits.

Position 2: Device physics

Job description: Carry out independent research on the physics of superconducting devices and circuits.

Position 3: Quantum materials

Job description: Conduct independent research on quantum materials with an emphasis on superconductors and superconducting hybrid structures.

Types of positions: Full professor, associate professor, assistant professor and postdoctoral researcher

Offer: Postdoctoral positions are offered salaries of up to RMB 150–200 thousand per year, depending on qualifications. The salary is negotiable for other positions. If suitably qualified, candidates can access further benefits through state, municipal and CAS talent recruitment programmes.

Requirements:

Doctoral degree in superconducting electronics, quantum electronics, condensed-matter physics or equivalent. Independent work experience in relevant research fields for two to five years or longer, depending on position;

- Papers published in high-impact journals in relevant fields;
- Two letters of recommendation;
- Research proposal (template available on request);
- Good team spirit;
- Experience of doing research overseas preferable.

Please send your resume and other application documents to:

Prof. Dr. Lixing YOU (lxyou@mail.sim.ac.cn)

Shanghai Center for SuperConductivity
Chinese Academy of Sciences
Website: sc2.sim.cas.cn

Q&A Chen Xiangmei and Zeng Xuan

Times are changing

Two women, 18 years apart, describe their experiences within China's research system. Chen Xiangmei (pictured, left) received her MD from Kitasato University, Tokyo, in 1986 and is now the director of the State Key Laboratory of Kidney Diseases of China at the PLA General Hospital in Beijing. Zeng Xuan (pictured, right) received her PhD in semiconductor physics from Fudan University, Shanghai, in 1997 and is now a professor at Fudan's School of Microelectronics.

In China's high schools, there are usually more girls studying languages and humanities than science and maths, and there is a stigma attached to the term 'female PhD' such that it has become a derogatory way of referring to an educated woman who is focusing on her career. A lot of parents tell their daughters that girls are not as good at science as boys. Have you ever faced this type of pressure?

Chen Treating women as if they are intellectually inferior to men has thousands of years of history in China's major ethnic groups, such as the Han. There are many families that hold onto this traditional line of thinking, especially in the less developed parts of the country. Often, such families would choose to support their sons to pursue a career

that requires more years of education. The situation is usually more balanced in large, economically dynamic cities and in educated families. The percentage of female students at top universities is increasing, and a substantial percentage of women are majoring in science or maths.

In 2001, I became a chief scientist funded by the 973 Programme, also known as the National Basic Research Programme, which is a prestigious funding scheme in China. I was the only female chief scientist of a 973 Programme in 2001. We need more women at the top level of society, not just in science, but also in the army, in the government and every aspect of society, otherwise we are wasting talent.

Zeng I was once that rare species: a female science PhD candidate! My interest in physics started when I was in high school in the 1980s. I was lucky that my parents never discouraged me from pursuing science. I read about the female physicist Xie Xide, who was president of Fudan University in Shanghai at the time. Scientists were like rock stars in the 1980s in China, and we were all familiar with the saying: 'Master math and science, then you have nothing to fear wherever you go'.

When I arrived as an undergraduate at Fudan University, my first choice was biomedical electronics but I was transferred to microelectronics. I was worried initially, but then I learnt that the microelectronics major

had been created by Xie, which inspired me to work really hard. I was fascinated from the first moment that I had access to computers and integrated circuits in the lab during my second year. I was fortunate to have many great mentors who nurtured my interest. No one discouraged me from studying science because I am a woman.

Are there the same opportunities for men and women to become scientists in China?

Chen After the People's Republic of China was founded in 1949, women in China were encouraged to work and were treated as equals to men legally, so I think there are the same opportunities for men and women to become scientists in China. But in real life, at least until now, women tend to spend more time raising children and taking care of their families.

Zeng Among the top high-school students, undergraduate students and graduate students in life sciences and medicine, women make up a similar percentage to men. That's less true in maths and physics, where men are still in the majority. So in many fields, the opportunity to become a scientist is equal for the sexes in China. At the very top, however, there are far fewer women. I think that is going to change as society becomes more aware of gender inequality (see 'The global picture of women in further education').

What is the availability of career breaks, maternity packages and flexible working for scientists in China?

Chen Nationwide, the minimum paid maternity leave is 14 weeks, but women who have babies later in life could have an additional 15 or 30 days depending on the region in which they work. Women scientists have the same maternity packages as women in other professions, but it can be a real challenge for them to catch up with everyone else when they return to work.

Zeng With regard to maternity, instead of providing a longer period of maternity leave, it's more important for a female scientist to have access to flexible working. An extended break can be detrimental to a scientist's career, but flexible working arrangements are still rare in China. Grandparents often step in to take care of the baby when a mother returns to work.

I have been a visiting scholar to several US universities and I was told that when female researchers are being evaluated for tenure, they are given two extra years in their evaluation period to compensate for time lost while raising children. Similar policies are also appearing in the national funding programmes in China, but there is not anything at the institutional level yet.

What about paternity leave — how common is it for men in China?

Chen It would be great if men in China would take a larger portion of the parental leave! I've

THE GLOBAL PICTURE OF WOMEN IN FURTHER EDUCATION

In the past two decades, China has boosted its share of women studying in tertiary education and, after overtaking its neighbour Japan in 2004, it is now in line with world averages.
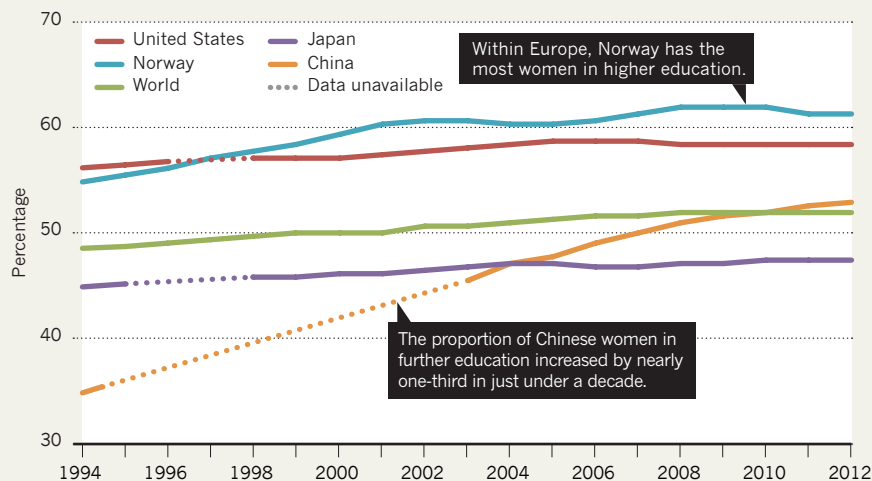
heard of men taking paternity leave to take care of their newborns, but I've never met anybody who has done so in real life. Traditional values still hold sway in China, so even if a man was willing to take time off to care for his baby, he would probably face social pressure from others. I hope that parents share the responsibility equally to take care of their children, but in reality, it is women who end up doing more.

"In many fields, the opportunity to become a scientist is equal for the sexes in China."

Zeng Men can take a few days of paid paternity leave immediately after the baby is born, but I, too, have never met any Chinese men who took substantial paternity leave. It seems that Chinese men are holding onto traditional values very tightly.

Are there biases in research evaluation that disadvantage women?

Chen In most of China's research-evaluation systems, women are treated exactly the same as men, but women who take time off to have children are often disadvantaged compared with men of the same age. So the National Natural Science Foundation of China youth programme changed its application age limit a few years ago to allow female scientists under 40 and male scientists under 35 to apply for its funding. This is a great way to mitigate the disadvantage faced by women.

Zeng I don't think there are any policy biases in research evaluation that actively disadvantage women, but family responsibility can sometimes slow a female scientist down. When I finished my PhD, there was a gap in professors of both sexes because the Cultural Revolution put higher education on hold for a decade and many new graduates chose to study abroad.

Fudan University's administration trained the young researchers to become lecturers and helped them to establish their own laboratories. I was lucky to take advantage of that.

What do you think would be the best way to encourage more women to pursue science?

Chen I think the best change would be to move away from traditional values. Young people with potential and who have good career prospects should receive support from their significant others and family regardless of their gender. Also, I tell all my students that if they want to become successful physicians, they should spend more time and energy on their work. It is also vital that they communicate effectively with their partners about what their jobs will require so they can share household duties. With support from their families, women can become outstanding medical scientists.

Zeng I am concerned that so few youngsters in China today aspire to be scientists. That is partly because teachers do so little to stimulate interest in maths and science, especially anything that directly appeals to girls. Students are more focused on simply getting good grades. The lack of encouragement at an early age may explain why there are very few women in maths and physical sciences in higher education. When I was studying geometry in high school, students who could draw neat lines to solve geometry problems were deemed smart and those who could not were deemed not smart. Years later, I met a mathematician in college and learned that the ability to draw such lines is not a good measure of ability: analytical geometry is better. We need teaching methods that will nurture girls' interest in maths and physical sciences. ∎

INTERVIEW BY PIAO LI

<div style="writing-mode: vertical">DALE EDWIN MURRAY</div>

RESEARCH IMPACT

A tale of two systems

Tensions between the old communist regime and modern market forces are hindering efforts to turn China's scientific discoveries into commercial advances.

BY PENG TIAN

It is definitely 'the worst of times' for Chu Jian. At the end of 2013, the engineer and former vice-president of the prestigious Zhejiang University in Hangzhou was arrested on suspicion of embezzling state-owned assets, and was formally charged in February 2015, just before Chinese new year. Chu was co-founder and director of Supcon Group, an engineering company based in Hangzhou in Zhejiang province. A major part of Supcon Group was Haina Software, a former subsidiary of Zhejiang Haina Science and Technology Company, a spin-off company from the state-owned university. Haina Software was sold to Supcon Group by 2003, and one of Chu's alleged crimes is transferring the better performing assets from Zhejiang Haina to Supcon, leaving poor-performing ones behind.

Chu's story is not unusual. Dozens of university officials and entrepreneurs are being swept up in China's ongoing anti-corruption efforts. But getting a clear picture of the extent of the corruption is hampered by difficulties in penetrating the murky rules surrounding the commercialization of state-funded research. Even as China is pushing to translate more basic research into societal benefits, this lack of transparency is preventing many would-be entrepreneurs from taking the plunge.

WESTERN INFLUENCE

Chinese universities and research institutes are reforming their research processes and evaluation systems to meet the higher standards of the international academic community and become more like Western research systems — particularly that of the United States. And to a certain extent these reforms are working: with better training and more resources, an increasing number of Chinese scientists are getting published in higher quality publications such as those covered by the *Nature Index* (www.natureindex.com).

The reform process is being fuelled by investment from local and central governments. China spent more than 1 trillion yuan (US$165 billion) on research and 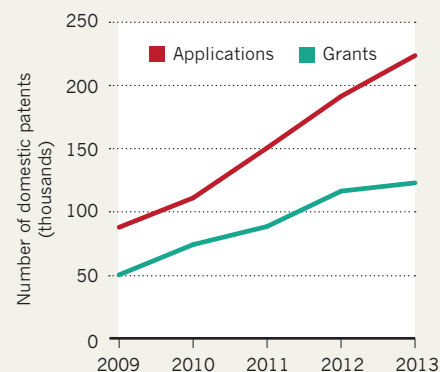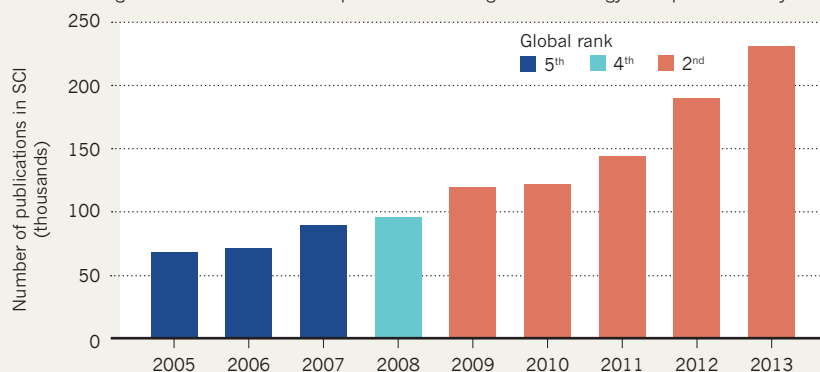development (R&D) across all sectors in 2013 — second only to the United States (see page S8) — and has the second largest output for academic publications and patents (see 'Rising output'). The hope is that basic research will feed into inventions and improvements for industry, and eventually boost the economy.

But when it comes to turning scientific discoveries into technological innovation, the results are disappointing. In the biomedical industry, for instance, most domestic drug companies produce either generic versions of existing drugs or drugs based on Chinese traditional medicines, which struggle to get into the regulated markets of developed countries. Even Xinhua, the official press agency of China, has criticized the technology transfer capabilities of the country's research system.

As the reforms of publicly owned research institutions continue, they throw up more uncertainty. "The atmosphere is changing in both universities and Chinese Academy of Sciences institutes — people can feel that things are coming," says Wang Liming, a 32-year-old neurobiologist at Zhejiang University. Wang, like many researchers of his

RISING OUTPUT

China's science and technology output continues to rise steadily, in terms of both numbers of publications in the science citation index (SCI) and of patent applications and grants. Yet there is no clear path for translating this technology to improve industry.

OBSTACLE COURSE

In response to criticisms of China's inability to commercialize its research, the government green-lit a 1-billion-yuan National Centre for Translational Medicine in October 2014. This centre is to be in Ruijin Hospital, which is affiliated with Shanghai Jiaotong University, and is expected to complete construction in 2017.

The Shanghai centre is the first in a series of projects for translational medicine from the government's National Science and Technology Infrastructure programme. There are four sister institutions planned, to be built in influential hospitals or medical colleges: two in Beijing, and one each in Xian in Shaanxi province and Chengdu in Sichuan province.

The main goal of the translational medicine centres is to more efficiently transform basic biomedical research into clinical applications, including new diagnostics, treatments and drugs. But many obstacles stand in the way. First is that although reform of China's scientific research system is under way, reform of its public hospital system is lagging behind. Around 90% of China's health care is provided in public hospitals. In theory, most public hospitals are wholly owned and financed by local and central governments, which strictly control staff numbers and salaries. But many hospitals have said that they do not receive enough funding and need to find additional resources. There have been reports of hospital staff allegedly accepting bribes from drug or device firms; one high-profile case involving British pharmaceutical giant GlaxoSmith-Kline, based in London, was exposed in 2013.

Recruitment is another issue. "Public hospitals are still in the old system, where there is less mobility of personnel," says Chen Zhiguo, a young neuroscientist at Xuanwu Hospital in Beijing. Hospitals such as the new Shanghai translational centre, which are affiliated with universities or medical colleges, may be able to get around this barrier by recruiting staff through their affiliates. But even these institutions often retain ties to the old system — and its policy and culture. Researchers at Xuanwu Hospital, including Chen Zhiguo, are obliged to publish a certain number of papers each year in journals covered by the Science Citation Index. This requirement takes up valuable research time and can stifle commercialization efforts. The result is a chicken-and-egg situation. "If a researcher successfully transfers her or his research to the market and gets economic benefits, the administration will loosen the requirement of publishing papers," Chen Zhiguo says. "But before that, you have to cope with it."

Then there are questions over financing. Because there is no Chinese equivalent of the United States National Institutes of Health (NIH), which funds clinical trials, researchers wishing to develop drugs must look elsewhere. "More than one-half of China's clinical trials are supported by drug companies," says Li Mengfeng, vice-president and dean of Zhongshan School of Medicine at Sun Yat-sen University in Guangdong. Some local governments and universities are trying to encourage clinical research that is initiated instead by independent investigators. One example is the 5010 Clinical Research Foundation, established in 2008 at Sun Yat-sen University. But there is nothing yet on the national level, says Li.

One attempt to bridge the funding gap is the Shanghai Guangci Translational Medical Research Development Foundation, which was established in September 2014 by Chen Saijuan, director of the translational centre, and her husband Chen Zhu, China's former health minister. The main goal of the foundation is to finance clinical research. The two have successfully worked together before, when they led a translational medical research project in China that resulted in a new treatment for leukaemia.

Even if all the obstacles could be overcome, these new translational centres are only one link in the 'bench to bedside' chain. "Every link of the chain has problems in policy and culture," says Li, who adds that China is still weaker than other developed countries in basic research, and that the administrative side of the scientific community hampers creativity. "Basic research is not driven by the realistic needs of clinical practice, it's about getting more impact factors." And, even if a scientist has a strong talent for innovation, the environment is not supportive, he contends. "Governments and institutes in China do not give scientists enough incentives to bring their research to application."

Successful translational research depends on many factors, says Xiao Ruiping, a physiologist and director of the Institute of Molecular Medicine at Peking University in Beijing. As well as providing incentives for researchers, other important elements to ensuring succes are improving laws and regulations concerning research commercialization, and ensuring there is availability of capital and appropriate industry infrastructures. "The problems inherent in China's translational medicine process cannot be resolved by a few scientists and projects," says Xiao. "How can these systematic defects be overcome?"

STATE OWNERSHIP CONUNDRUM

Other countries have been through the process of promoting commercialization of the fruits of publicly funded research. The key piece of legislation in the United States is the Bayh–Dole Act of 1980. Prior to this law, any inventions derived from federally funded research belonged to the government. Bayh–Dole gave those universities, institutes and companies

"Chinese scientists want to be rich and famous."

generation, has entrepreneurial aims. But he acknowledges that in these unsettled times, many researchers will cling even more tightly to the old system. "Getting a publication in journals with a high impact factor becomes the priority for scientists when they feel unsafe," he says, "because they know it will probably help them survive any kind of top-down evaluation led by the administration."

who received federal grants the right to profit commercially from their inventions.

China has introduced similar polices, and has solidified them into formal rules in the Science and Technology Progress Law of 2007, but has had less success than the United States at fostering the commercialization of university invention. The reason lies in the different ways the two countries treat their universities and research institutes — in the United States these organizations have more independence from the government than their Chinese equivalents, and also have autonomy as market participants. China's universities and institutes, are all controlled by the government. Any economic activities must be approved by the administration of state-owned assets, which can hinder attempts by would-be entrepreneurs to commercialize their research.

Such a situation is facing Chen Zhiguo, who was hoping to transform his neural-stem-cell research into a clinical treatment when he returned to China in 2012 having completed a postdoc at Stanford University in California. Two years on, the results from Chen Zhiguo's animal experiments are promising, but because his research is funded by the government the project will need additional capital to move into clinical trials. Securing capital means transferring intellectual property, which is controlled by the hospital, which in turn is controlled by the state. "There are many difficulties," Chen Zhiguo laments. "How can we balance the interests of the hospital, the start-up and the researchers?"

The biggest problems are that current regulations concerning technology transfer are unclear and authority oversight is unpredictable. A young university researcher, who wishes to remain anonymous for fear of repercussion from his university administration, explains: "If I want to transfer my research achievements to the pharmaceutical industry, my only choice is to cross a 'grey zone' when it comes to intellectual property regulations. There is no clear legal track, so infractions are inevitable." It is impossible, he continues, for a scientist–entrepreneur to avoid these traps forever, as Supcon's Chu found out. "Faults could be investigated by the government or prosecutors in the future, particularly if a company becomes big and successful." Consequently, many inventors simply avoid the maze of policies and regulations. If they are unwilling to transfer their innovations through 'grey zone' agreements, their only choice is to abandon their inventions altogether.

There is plenty of entrepreneurial spirit in China. "Chinese scientists want to be rich and famous," says Xiao, who wants a "clear and stable system" that enables start-ups to flourish. The research outputs of Xiao's laboratory belong to Peking University, whose Office of Science and Technology Development is responsible for transferring her research to industry. "The

Chu Jian in 2006, who was then the vice-president of Zhejiang University in Hangzhou.

office is a part of the administration department of Peking University and does not have its own intellectual property lawyer," she says. Xiao worked for many years at NIH, and contrasts the situation in China with that of the United States. "I had nothing to worry about in terms of commercializing my research there. The NIH has a strong patent office that takes care of intellectual property matters. Scientists in China have to deal with things by themselves, and the patents are often fragmented."

The research community in China wants the government to acknowledge the contribution made by scientists working on publicly funded projects and to give them stronger incentives to carry out research with potential commercial impact. For research scientists, ownership priority over their inventions is an important step in that direction. Speaking at the Research Assessment and Evaluation symposium in Shanghai in October 2014 (see page S10), Shen Wenqing, former deputy director of the National Natural Science Foundation of China, said: "The most important reform is reducing the administrative intervention from the government." At the local level, some provinces are trying small-scale reforms: for instance, since 2014, a research team in Hubei province can now own up to a 99% share of the benefits of their research. But this applies only to projects funded by the provincial government of Hubei — those funded by the central government allow for no such ownership.

PARADOX OF TRANSITION
The shadow of the old system still hangs over every corner of China's scientific enterprise. The balance of administrative power of governments and of market forces is a frustrating paradox.

Entrepreneurs may find themselves navigating a complicated sea of regulations issued by different government departments, in many cases trying to get one department to intervene against the other. "The reality is that when you need them, they disappear and you cannot find them," the young researcher says, "and they always appear where they should not be." For example, the China Food and Drug Administration (CFDA) has failed to remove traditional Chinese medicine injections from the market, even though their safety and efficacy are unproven and people spend billions of yuan on them each year. Yet the CFDA requires start-ups to have their own production factory and certificate of Good Manufacturing Practice, which makes it almost impossible for an innovative pharmaceutical company to get up and running. Government officials are trying to intervene too much in market processes, explains Guo Chongqing, a research engineer at Tongji University in Shanghai. "Officials in governments do not have crystal balls. How could they plan the way that industries will develop?"

The downstream businesses that are essential in supporting the translational medicine ecosystem are also unformed. Most venture capital firms in China are reluctant to consider the large risks associated with drug development — and often lack the expertise to understand the technology, says Tang Yanmin, director of the Beijing office of Morningside Ventures, a venture capitalist headquartered in Hong Kong and one of the few firms that invest in early stage biotechnology in China. Tang adds that Chinese venture capitalists generally prefer the safer returns from investing in generic drugs and medical devices, based on cost advantage and state protection.

"How can we balance the interests of the hospital, the start-up and the researchers?"

Tang believes that the returns for innovative drugs will rise in the future as domestic pharmaceutical companies complete more original R&D and respond to the government's emphasis on advancing the industry. The fruits of China's burgeoning research are ripe for harvest. But at the moment, with so little clarity on how to proceed, concerns about corruption charges — such as those facing Chu — are too serious to ignore. "Hoping that the governments will demolish all the invisible walls and draw clear lines overnight is unrealistic," says Zhejiang's Wang. He understands why so many scientists who tried to commercialize their research gave up, but he is determined to succeed. "Yes, there are still many obstacles from the old system," says Wang. "But I want to get things done." ∎

Peng Tian *is a freelance journalist based in Beijing.*

Zhejiang University
SEEKING THE TRUTH AND PIONEERING NEW TRAILS

Zhejiang University (ZJU), one of the leading research-intensive universities in China, is located in the beautiful city of Hangzhou. Research at ZJU spans 12 academic disciplines: philosophy, economics, law, education, literature, history, art, science, engineering, agriculture, medicine and management. Of the 22 subjects listed in Thomson Reuters Essential Science Indicators, ZJU is rated in the top 1% in the world for 16.

As a national university, ZJU is building itself into an innovative and comprehensive research institution of world-class standing. ZJU seeks the truth and pioneers new trails. It is dedicated to creating and disseminating knowledge, advancing civilization, serving and leading society, and promoting national prosperity, social development and human progress.

ZJU sincerely invites talented researchers to join it so that together we can create a glorious future.

Contact:
Talents Office, Human Resources Department
Zijingang Campus, Zhejiang University
No. 866, Yuhangtang Road, Hangzhou
Zhejiang, 310058, P. R. China

Tel: +86-571-88981390, 88981345, 88981656, 88981453 (Postdoc)
Fax: +86-571-88981976
E-mail: tr@zju.edu.cn
http://talent.zju.edu.cn

Why did you choose Zhejiang University?

"Zhejiang University acts as a bridge to the exciting and fast-growing Chinese market. It allows researchers to harvest real-life problems and provides abundant opportunities for applying research outcomes to real systems. Besides providing opportunities to work with the best students in China, Zhejiang also has an excellent research environment that nurtures world-class research."
Professor Wenyuan Xu, College of Electrical Engineering

"For a young researcher, China is a fascinating country to begin an independent career. The creativity and pace of modernization will impress everyone who visits, and having a position here gives you the opportunity to focus on what really matters: research and the education of a new generation of scientists. Zhejiang University is one of the top universities in China and provides starting faculty with excellent working conditions.

The Department of Chemistry supports its young professors wherever it can, gives them much freedom and welcomes interdisciplinary collaboration. I am happy to have brilliant and helpful colleagues as well as smart, hard-working students."
Research Professor Simon Duttwyler, Department of Chemistry

"As a former PhD student at Zhejiang University, I have always valued its strong support network, so after doing two postdoc programmes here, when the opportunity came for me to stay and teach, I took it without hesitation. I am proud to be on the faculty of the College of Agriculture and Biotechnology and I feel like I can never repay my alma mater for transforming me from a young boy into the quality teacher that I am today."
Lecturer Imran Haider Shamsi (Ying Lan), Department of Agronomy

"I decided to return to China in 2009 after working abroad for 15 years as I believe that China is the best place to find sustainable development solutions to address our planet's limited resources. Indeed, the Yangtze River Delta, where Zhejiang University is located, is an area that has prospered continuously for thousands of years without major environmental decay. In particular, I enjoy the freedom to pursue my area of interest as a researcher and the open-minded atmosphere at Zhejiang University."
Professor Xiaogang Peng, Department of Chemistry

"Although I officially joined Zhejiang University in 2011, I have been collaborating with researchers at Zhejiang University since 2002. I have found it to be a place with great academic freedom, and I believe that's the reason why Zhejiang University is one the best universities in China. In addition, I am grateful to be able to work in the beautiful city of Hangzhou, where I grew up. Zhejiang University is also strong in engineering, which provides a fantastic platform for physical science research and interdisciplinary collaboration — activities that are of particular importance to me."
Professor Fuchun Zhang, Department of Physics

"Zhejiang University's location in the economically dynamic region of the Yangtze River Delta provided much opportunity for translational research and a large patient base for biomedical research."
Professor Lanjuan Li, Director of the State Key Laboratory for Infectious Diseases Diagnosis and Treatment

浙江大学
ZHEJIANG UNIVERSITY

:insideview
profile feature

Lanjuan Li, *Director and Chief Scientist of the CCID*

Lanjuan Li is the director and chief scientist of the CCID as well as an academician at the Chinese Academy of Engineering, a professor of Zhejiang University and the director of the State Key Laboratory for Diagnosis and Treatment of Infectious Disease.

CCID was founded in May 2012 as a non-profit research centre with the goal of promoting both basic and clinical research into infectious diseases. It was established through the collaborative efforts of five institutions in mainland China and Hong Kong: the Collaborative Innovation Program of the Ministry of Education of China, the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases of Zhejiang University, the State Key Laboratory for Infectious Disease Prevention and Control of the Chinese Center for Disease Control and Prevention, the State Key Laboratory for Emerging Infectious Diseases of Hong Kong University, and Beijing National Engineering Research Center for Biochip Technology of Tsinghua University.

Q: Why is a multidisciplinary approach important?
CCID has five main research focuses: basic and clinical research of emerging infectious diseases; the development of an alert system and an emergency response system for major infectious disease; research into mechanisms of how acute viral hepatitis turns chronic and severe; the treatment of severe liver failure using the combination of artificial liver and liver transplantation; and investigation of the relationship between infectious disease and human gut microecology.

Q: Why is a multidisciplinary approach important?
Since infectious diseases have complex pathogens and their transmission often involves multiple regions and species, their diagnosis and treatment often require multidisciplinary expertise. By establishing a collaborative innovation centre among several research institutions, we can better study the cycle behind infectious disease development and the interaction between the body's immune system and pathogens. We need the ability to rapidly source the pathogens of emerging infectious diseases and establish an effective early warning system and a systematic approach to diagnosis and treatment.

Q: What work are you doing in the area of liver failure?
My research team developed a hybrid artificial liver support system that provides plasma exchange, hemoperfusion, hemofiltration and continuous hemodiafiltration that can

be adjusted to the conditions of an individual patient. This system is now widely used in China, and we are continuing to improve it.

Q: H7N9 avian flu continues to receive a lot of attention internationally. How has CCID contributed to the research of H7N9?
Of all the provinces in China, Zhejiang province experienced the largest number of H7N9 cases since its initial outbreak in March 2013, and many of patients were treated at the First Affiliated Hospital of Zhejiang University. Together with our collaborators, we published a clinical analysis and characterization of viral genome in the journal Lancet in April 2013 and clinical findings for 111 cases of H7N9 in the New England Journal of Medicine in May 2013. As soon as we realized that H7N9 was being transmitted from live poultry to humans, we advised the government to close live poultry markets. In addition, we have studied the genetic mutation and cytokine storm of H7N9 and a developed vaccine against the virus.

Q: What were the findings of research into the changes of the human gut microbiome in liver cirrhosis patients?
Previous studies had suggested an association between liver diseases such as cirrhosis and changes in the composition of the human gut microbiome, but definitive associations between gut microbes and afflictions of the liver were lacking. We performed a microbiome-wide association study of stool samples from 98 liver cirrhosis patients and 83 healthy controls of Chinese ethnicity.

Quantitative metagenomics analysis revealed 75,245 genes that were significantly different in abundance between the cirrhosis patients and the control group. Many of these genes could be grouped into 66 clusters that represent cognate bacterial species. Of these 66 clusters, 28 species were enriched in liver cirrhosis patients and most were of buccal origin (mostly Veillonella and streptococci). Based on this analysis, we identified 15 genes that could form the basis of a highly accurate discrimination index for diagnosis.

Q: What are CCID's future plans?
CCID has five main active research areas, these include: (1) Basic and clinical research of emerging infectious disease, (2) Developing an alert system for major infectious disease and an emergency response system, (3) Clarifying the mechanisms behind how acute viral hepatitis becomes chronic and severe, (4) Combining artificial livers and liver transplantation to treat severe liver failure and (5) Examining the relationship between infectious disease and human gut microecology. CCID is using metagenomics and metatranscriptomics tools to clarify the role of microorganisms in the progression of liver diseases. CCID is dedicated to finding fresh strategies for diagnosing, treating and preventing diseases such as liver cirrhosis. We are looking into clarifying the effects of antibiotics and probiotics on the liver, to explain the effects and mechanism of probiotics like lactulose on disease progression. We are also studying the alteration of gut microbiomes of liver failure patients in the hope of finding new candidate probiotics for liver disease treatment. We are seeking to expand our international collaborations. Members of CCID will be attending more international conferences to communicate and cooperate with researchers in different countries. We are hoping to draw on a wider range of expertise and aim to become the Chinese equivalent of the National Institute of Allergy and Infectious Diseases in the United States and the Pasteur Institute in France. We aim to significantly reduce the incidence of infectious diseases and the associated mortality rate in China.

COLLABORATIVE INNOVATION CENTER FOR DIAGNOSIS AND TREATMENT OF INFECTIOUS DISEASES

In 2011, the Chinese Government introduced an initiative to promote cooperation between leading research institutions in China and abroad with a view to boosting the innovation capacity and accelerating the internationalization of China's higher education.

Zhejiang University was awarded substantial funding from this initiative, which it used to form the Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases (CCID) in May 2012 in conjunction with Tsinghua University, the University of Hong Kong and the China Center for Disease Control. Lanjuan Li, the director of the centre is joined by senior principal investigators Jing Chen, Kwok-Yung Yuen, Shusen Zheng and Jianguo Xu, associated with the collaborating institutions listed above. In 2014, the Chinese Government made CCID the only national-level collaborative innovation centre for medical research.

The five missions of CCID

CCID seeks to research, prevent, diagnose and treat infectious diseases in China with the goal of creating solutions that have the greatest possible impact. The centre is currently focusing on the early warning, prevention, pathogenesis, molecular mechanisms, diagnosis and treatment of infectious diseases — especially influenza and hepatitis. It has research platforms for researching infectious pathogens and providing warning and forecasting as well as conducting pathogenesis, diagnosis,

treatment and bioinformatics. CCID also has the capacity to clinically evaluate new drugs for infectious diseases and perform animal studies and clinical diagnosis and treatment.

CCID's first mission is to establish effective early warning and monitoring systems for new infectious diseases, detect trace pathogens of emerging infectious diseases, determine the mechanisms responsible for the sudden onset and subsequent intensification of diseases. In addition, it involves developing clinical treatment strategies and treatments and reducing the high mortality rates of patients with rapid disease progression.

The centre's second mission is to utilize community disease-prevention programmes and epidemiological data cloud platforms to monitor the occurrence of emerging infectious diseases. CCID is also investigating the patterns behind the spread of disease with the aim of developing forecasting and early warning systems. CCID is dedicated to enhancing the capabilities for rapidly identifying and preventing infectious diseases in China.

CCID's third mission is to study how viral hepatitis and other major infectious diseases become chronic and severe and can eventually lead to cancer. The centre also aims to establish early warning and forecast systems for chronic hepatitis B patients.

To tackle the problems of severe liver failure and the high mortality rate of liver disease, CCID is striving to clarify the mechanism of liver failure induced by severe liver

disease, conduct effective and individualized artificial-liver treatment and research liver transplantation to significantly reduce mortality. This is its fourth mission.

The fifth mission for CCID is to systematically characterize the key development stages, structure and function of human gut microflora of patients suffering from H7N9 influenza and other emerging infectious diseases, as well as liver disease, tuberculosis and other major infectious diseases. In particular, CCID will investigate how the relations between the microecology, the host's immune system and metabolic changes in the hope of achieving a breakthrough in identifying gut microbiome dysbiosis.

Looking for international collaboration and talented researchers

Under the leadership of academician Lanjuan Li, CCID has a formed a pool of talented researchers with clinical and basic research experience across many disciplines. It has already published 22 papers in journals with impact factors exceeding 30, such as *Nature*, *Lancet* and *the New England Journal of Medicine* and holds over 100 patents. CCID responded rapidly to the first wave of avian influenza A H7N9 and discovered novel liver cancer biomarkers. CCID is actively looking to recruit talented researchers from overseas and establish more international collaborations.

Q&A Anthony Cheetham
China still rising

Materials scientist Anthony Cheetham, vice-president of Britain's Royal Society, outlines how China's approach to recruiting and funding scientists is driving steep increases in Chinese research productivity.

What are your experiences of working with researchers in China?

I have been to China about 15 times in various capacities. As a science adviser to the board of Unilever from 2000 to 2008, I visited the company's lab in Shanghai at least once a year. From 2004, I ran the International Centre for Materials Research at the University of California, Santa Barbara; it was clear that science was really taking off in China, so building links there was one of my top priorities. I went on to co-chair the first US–China workshop on advanced materials, for which we took a delegation of faculty members, students and postdocs to Beijing. After I moved to Cambridge in the United Kingdom, we continued the link with a UK–China workshop over here.

In which types of institution does the best scientific research take place?

Much of the leading science is done in the institutes of the Chinese Academy of Sciences (CAS), which subsumes two different roles. It has academicians, like those at the Royal Society or other science academies, but it also runs roughly 100 institutes covering all facets of science and engineering. These are distributed around the country with something of a concentration in Beijing and Shanghai. It is a little bit like the German system in which a lot of the outstanding research takes place in Max Planck institutes rather than in universities. Although most of the best science happens at CAS, China's top universities are also very impressive.

In which fields is the country strongest?

Overall, the number of Chinese publications in leading journals is roughly comparable to that of the United Kingdom, France, Germany or Japan. What was striking was that when I looked at the trend from 2000 to 2014, although there were variations from year to year, there was steep upward growth in all areas for China.

Among the major science categories, Chinese growth was steepest in chemistry, to the point where the country is now second in the world in terms of overall output behind the United States. In physics, China is comparable with the United Kingdom, perhaps slightly overtaking us, yet in life sciences it still lags behind — although again the gap is narrowing.

JON ENOCH

My interpretation is that China has invested heavily in science across the board, but it is easier to make rapid progress in a discipline like chemistry because you do not need the large, sophisticated infrastructure that you often do in physics. And, on the other hand, funding for the life sciences is particularly strong in Britain.

Is the funding application system fair and effective?
In 2010, I went to China as part of the first international evaluation of the National Natural Science Foundation of China (NSFC), which is their equivalent of the US National Science Foundation and of bodies like the UK Engineering and Physical Sciences Research Council (EPSRC). I was impressed by the quality of the NSFC's processes when it came to funding grants. We were a small committee of half a dozen international scientists and two or three Chinese scientists, and we were given unfettered access to anybody we wanted to speak to and given a large amount of data about trends in applications, about their volumes and pass rates. We were able to interview stakeholders in the whole process, including not only people from the NSFC, CAS and China's Ministry of Science and Technology,

but also administrators and grant holders from universities. We found that the NSFC commands a huge amount of respect among Chinese scientists.

Funding that goes through CAS is done differently, but nevertheless Bai Chunli, its president, is looking at ways to improve the evaluation of the performance of its institutes and individuals within them. CAS feels it has to work a bit harder to justify its large tranche of funding from the public purse. It has been seeking international advice on this, and the Royal Society sent a delegation over there a couple of years ago as part of that process.

What specifically does the NSFC do best?
It is doing peer review very well. The programme officers we met were outstandingly good. There was a lot of transparency in the system so that even when people did not get funding, they felt they had been given a fair opportunity. They are also very proactive on scientific misconduct and have an established system for monitoring it and prosecuting cases. To be honest, if we had done the same in-depth evaluation of a funding body like the EPSRC, I suspect that the council would have come out of it with a similar or maybe even lower score than the NSFC.

> *"Young scientists in China have incredible levels of passion and drive that we don't often see in the West."*

Do Chinese scientists face particular pressures relating to how their work is disseminated?
Salaries and funding decisions are directly related to numbers of publications in high-profile journals. This emphasis on quantitative metrics also applies to the evaluation of candidates for academic positions. I know from supervising Chinese postdocs that if they want to go back to China, they need a minimum of four or five publications in a well-defined list of high-quality journals, and that they must be either the first or a corresponding author.

It is a competitive system and the quality of science coming out of China is getting better each year. I personally would like to see them placing greater emphasis on evaluating people and reading what a researcher has done, rather than ticking boxes and using the impact factors of journals as a surrogate for quality. I suspect that the number of people applying for positions is huge. If you are deluged with candidates, that might explain some of the rather crude tools used to filter them.

Critics suggest that this focus on metrics encourages a culture of seeking quick, short-term wins. Do they have a point?
There are different ways of responding to the pressures in the Chinese system. Certain things can't be done quickly. There are people

everywhere who are interested in short-term success and I don't think the Chinese are any different in that respect.

I do think that the emphasis on publications in high-impact journals can drive some bad behaviour. There is concern that in a minority of cases it can encourage falsification of data and plagiarism. These things happen everywhere but are perhaps more likely when rewards are tightly linked to outputs.

Compared to other countries, China spends proportionally less money on basic science and more on developing existing technologies. Does this make the country less likely to make fundamental discoveries?
Based on the sheer volume of work its scientists publish in world-leading journals, it is ill-informed to think that China is not very good at discovery. These are not journals that publish derivative stuff. There may be more money going into developmental science, but the amount in absolute terms going into discovery science is probably higher than in most Western countries apart from the United States. The processes they have in place clearly enable them to support an ever-growing amount of high-quality science.

How good are Chinese scientists at translating their scientific discoveries into applications that benefit society?
China has a long way to go when it comes to capitalizing on its scientific base. The concept of gaining real economic advantages from its investment in science is more embryonic there than elsewhere. In the United Kingdom we have the Catapult innovation centres to help to do that, and other countries across Europe have similar institutions. Whether it is training in entrepreneurship for scientists or helping them to start companies, these things are not as well developed in China. Venture capital exists but it is nothing like the level in Britain, let alone the United States. The Chinese recognize that their capability in science has outstripped their ability to take advantage of it, and they need to build up their capabilities.

What could outsiders learn from the way Chinese researchers manage their science?
We could learn a lot just by observing the speed with which they have gone from being minor league players 20 years ago to being in the very front ranks of international science today. It's a phenomenal success story. Beyond the substantial and sustained financial investment, what I can say from my experiences of taking Western students to workshops in Beijing is that they are generally amazed by the levels of enthusiasm of their Chinese counterparts. There is no doubt that young scientists in China have incredible levels of passion and drive that we don't often see in the West.

INTERVIEW BY NIC FLEMING

PERSPECTIVE

Give youth a chance

Despite the growth in science spending and research output in China, young researchers are struggling to prosper, says **Chuan–Chao Wang**.

Given the decades of continued growth in spending on research and development and huge investments in buildings and equipment, it is no wonder that Chinese scientists, officials and the general public are expecting great advances in science and technology. Yet the reality is that an increasing number of young researchers, including graduate students and postdoctorates, are choosing to escape from research — a situation that has become a topic of heated national debate.

Many recruitment notices for senior research positions require a candidate to have a good publication record, impact factor and citation metrics. Young scientists strive to have a paper published in *Cell*, *Nature* or *Science* — elite, peer-reviewed journals known collectively in China as CNS. A paper in a CNS journal is an unwritten requirement for a better salary or a promotion — especially in China's top universities and institutes. A CNS paper could change a young researcher's academic career — or even his or her life.

Focusing evaluation on such inflexible terms leads us, as scientists at the beginning of our careers, to adopt utilitarian rather than academic motives. Rather than follow our natural curiosity for research, we choose small, straightforward projects so that we can obtain results quickly. We need a flow of publications to ensure continued financial support. But every researcher also wants the opportunity to make a CNS splash.

HOW TO RETAIN TALENTED SCIENTISTS

The irony is that although investment in science is rising in China across the disciplines, there are few funds or fellowships for newly graduated PhDs. The largest is the Young Research Fund, which represents about 20% of the budget of the National Natural Science Foundation of China (NSFC). However, each project receives only 240,000 yuan (US$38,000) over three years. The second option is a grant from the China Postdoctoral Science Foundation, in which a successful applicant can secure 50,000–80,000 yuan for two years' work. Neither amount is sufficient to purchase the necessary laboratory reagents and consumables to conduct research. For this reason, many young researchers apply for overseas postdoctoral positions, giving some of their best years of research to other countries — and often staying there.

As a consequence — and in contrast to most Western countries — research projects in China rely on graduate students rather than postdocs. Exacerbating this trend, many Chinese universities and institutes hire as faculty only those candidates who have overseas research experience. China is, in effect, outsourcing the training of its young researchers. But the reality is that most scientists who graduate with a PhD from a Chinese university will not receive overseas offers; their choice is to stay at home and find work in a domestic laboratory and be paid a low salary with little hope of promotion, or give up their academic career entirely to seek alternative employment.

Thirty years ago, Ray Wu, a molecular biologist at Cornell University

in Ithaca, New York, initiated the China–United States Biology and Biochemistry Examinations and Applications (CUSBEA) programme, which allowed for the first time large numbers of young graduates from China to pursue PhDs at leading universities in the United States. Many CUSBEA students have since become leading biologists in the United States as well as in China. Following Wu's death in 2008, the Ray Wu Memorial Fund established a prize in his honour to inspire and reward PhD students who wish to further their education at universities in China, Hong Kong, Taiwan or Singapore. Unlike China's typical evaluation system, the Ray Wu Prize — an award of US$3,000 plus the opportunity to apply for up to $5,000 to attend international conferences — does not focus on a student's history of journal publications, overseas experience or examination results.

Instead, the prizewinners, of which there could be ten or more each year, are determined by peer review by a selection committee consisting of eminent Chinese biologists who assess each candidate's capacity for creative ideas, independent thinking and dedication to his or her field of interest. Applicants submit a statement describing why they chose to study science, the academic projects they would like to pursue at university, and their intended future career. They also have a face-to-face interview with the committee. In the years, or even decades, after the prizes have been awarded, the committee remains available for advice as the researchers develop their scientific careers.

The Ray Wu Prize is a comprehensive and impartial evaluation process that also offers advice and support. Moreover, it is starting to be recognized by Chinese universities and institutes as a factor in faculty appointments. In 2014, a scientist who was awarded the Ray Wu Prize was appointed as assistant professor (tenure-track) at ShanghaiTech University, despite not having overseas research experience.

For decades, the minds of Chinese students have been shaped by passive rote education. But scientific progress requires not recitation of known concepts, but new ideas. The capacity to break the mould and truly innovate — the features that the Ray Wu Prize emphasizes — are crucial for the Chinese scientific community to deliver the results that the government and the people expect from their investment. The potential of new graduates cannot be fully reflected by publication metrics or overseas experience and as a consequence, China is currently losing their contribution. China must invest more in its promising young scientists, and in more effective ways of measuring their achievements. ∎

> ALTHOUGH **INVESTMENT** IN SCIENCE IS **RISING** IN CHINA, THERE ARE **FEW FUNDS** FOR NEWLY GRADUATED PHDS.

Chuan-Chao Wang *is a PhD candidate in biological anthropology at Fudan University in Shanghai. In June 2015, he will become a postdoctoral researcher at the Max Planck Institute for the Science of Human History in Jena, Germany. He won the Ray Wu Prize in 2012.* E-mail: cchao.wang@gmail.com

Q&A Jin Dong-Yan
The global view

Jin Dong-Yan is a virologist in the University of Hong Kong's department of biochemistry. With a bachelor's degree from Sun Yat-sen University in Guangzhou, a PhD from the Chinese Academy of Preventive Medicine in Beijing and a postdoctoral stint at the US National Institutes of Health, Jin benefits from the experience of working in vastly different research systems.

Why did you decide to go to the United States?
Twenty years ago, China had well-equipped facilities, especially in my field at my institution. The infrastructure to support fundamental research in biomedical sciences, however, was not well developed. I decided to do my postdoctoral research in the United States because the country was more advanced and had more world-class research opportunities than China.

What brought you back to Hong Kong?
First, I am Cantonese, so Hong Kong was a natural choice for me and my family. Second, when I joined the University of Hong Kong (HKU) in 1998 [a year after it officially reverted to Chinese rule], the city had already developed an advanced infrastructure to support basic research. I believed that Hong Kong was the best place to do science in Asia, and that I would do well. Third, HKU approached me with an attractive offer.

What are some of the strengths and weaknesses of Hong Kong's research system?
The greatest strength is international peer review. All proposals submitted to Hong Kong's Research Grants Council are subjected to a rigorous review process handled by a panel comprising world-renowned, independent experts. Unfortunately, because of limited resources, researchers in Hong Kong seldom have the chance to propose big projects. They have to be realistic and their projects tend to be more conservative.

What is the tenure system at HKU like?
HKU did not have a well-developed tenure system before I joined, but it was in the process of setting one up. Before that, people would be tenured immediately when they joined the university. That is no longer the case: today, people are given tenure only after the satisfactory completion of two three-year contracts, and there is no automatic promotion. In this respect, the HKU tenure system has become more similar to that at universities in North America. Beyond awarding tenure, we also have a system of performance review and development to evaluate our faculty annually. For non-clinical staff, we look at teaching, research and community service; clinical staff are also evaluated by their clinical service.

How does the university decide who to recruit?
We look at publications, but that is just one of several important factors. For example, we look at whether the candidate's areas of expertise complement what we already have. And during an interview, we evaluate communication and interpersonal skills. We need to find out whether this candidate is the one we really want and will enhance us in the years to come.

What is the system like in mainland China?
The mainland system has evolved over the years. When I left Sun Yat-sen University 30 years ago, performance reviews for tenure and promotion were run by local researchers, who might be influenced by personal connections. After many years of criticism, the system became more modern and international. One of the improvements was to weigh journals' impact factors when assessing the value of a researcher's publications. This encouraged faculty members to publish papers in top international journals. However, the system went too far and placed too much emphasis on impact factors and not enough on international peer review. The abuse of impact factors is out of proportion now, and that is the main weakness of the current system in mainland China.

What brought about these changes in the Chinese system?
The main driver was that many Chinese scientists based in the United States criticized the Chinese system and suggested a change from rule-by-people to rule-by-merit. They found that funding and faculty promotion are decided in the US in a merit-based way, which they felt is superior to the system in China. The Chinese government finally took some of their advice. Unfortunately, the Chinese evaluation system has gradually evolved into not a merit-based but a metrics-based one.

Is Hong Kong's research system still better than mainland China's?
Yes, our system is more advanced and fairer because we are much closer to the international standard. Also, on the mainland there is still too much interference in research by the government. That is not what scientists need when they are trying to do good research. Mainland China is making progress to improve its system, but there is still miles to go.

INTERVIEW BY FELIX CHEUNG

POLYU: WHERE INNOVATION MEETS APPLICATION
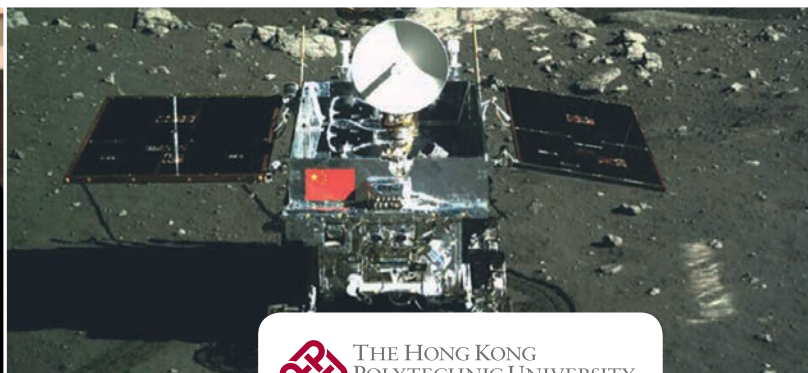
The Hong Kong Polytechnic University (PolyU) firmly believes that research is an integral part of academic life on campus. We are highly cognizant of the role technology plays in a knowledge-based economy and its growing importance in enhancing our well-being, economic growth and global competitiveness. The technological breakthroughs fostered through our research initiatives have been pushing the boundaries of sustainable development and changing our world for the better.

Over 2,700 research projects were conducted in 2013–2014. These ranged from 'down-to-earth' projects that touch people's daily lives to 'up-to-space' innovations that are expected to have far-reaching effects on the development of humankind. This research received funding of HK$1,541 million and involved more than 1,000 academic staff and about 1,000 research personnel.

Advancing knowledge and pushing technological frontiers

Over the past few years, PolyU's research innovations have continued to break new ground and contribute to technological advances that benefit the world.

Aviation Services Research Centre

Since the opening of its first phase in 2013, the Aviation Services Research Centre (ASRC) has secured more than HK$70 million in funding from the Innovation and Technology Fund as well as the participation of The Boeing Company, Hong Kong Aircraft Engineering Company Ltd and Hong Kong Aero Engine Services Ltd.

The ASRC has now embarked on a number of aviation services research projects that will address the pressing needs of the maintenance, repair and overhaul industry. The first project, Laser Projected Drilling Templates and Robotic Drilling, will design a process using a portable laser drilling template that will save manpower and streamline the workflow of replacing parts on the outer skin of an aircraft.

The second project, Mechanical Refurbishment of Aviation Parts, will develop advanced methods for mechanically refurbishing aircraft parts that will greatly cut maintenance costs. This will be achieved by developing a multi-axis probing technology and the world's first mechanized method for re moving distortions from turbine backing ring seals. Another ten projects are currently being undertaken.

Railway technologies gain global recognition

Interdisciplinary researchers from PolyU's Faculty of Engineering have been contributing to railway technology advancement through optical fibre and ultrasonic wave sensing technologies for monitoring the structural health and safety of the nation's rapidly expanding high-speed rail.

Recently, the team's efforts were recognized by Berthold Leibinger Stiftung, a German organization that promotes science research. The organization confers its coveted Innovationspreis awards every other year to researchers who have made important advances in the use of lasers. In 2014, Michael Liu Shun-yee, Tam Hwa-yaw and Ho Siu-lau from the Photonics Research Centre and the Department of Electrical Engineering were named second runner-up for adapting fibre Bragg grating (FBG) technology for use in railway monitoring systems.

Using light to monitor train conditions

The new FBG sensors, which were developed at PolyU's Specialty Optical Fibre Fabrication Laboratory, are specially constructed to both detect and transmit information about a wide variety of problems. FBGs are built inside a fibre-optic cable. Their design allows them to reflect laser light of certain frequencies back to its origin while letting other frequencies pass. The new sensors take advantage of the fact that any movement in the fibre-optic cable changes the shape of the sensor inside it, which in turn alters the frequency of the reflected light. A computer can analyse these changes in the light to determine the type and degree of movement in the cable. The team initially installed their new sensors on the tracks at Hong Kong's train stations with the sole purpose of counting the number of trains entering and leaving. Vibrations on the rails shake the sensors, deforming them enough to send signals to the monitors. While the FBG sensors performed this job very well, the researchers soon discovered more subtle uses for them.

This technology has many advantages over electronic systems. It is immune to the electromagnetic interference and lightning that can affect aboveground electric circuits, and the signals can be detected over distances as long as 100 kilometres. Furthermore, the sensors are passive and thus do not require power supplies. The system only needs a laser to generate light and a spectroscope to analyse the returning signals. Finally, the research team estimates that hundreds of sensors can be integrated into a single length of fibre.

The equipment is also physically robust. The fibres are made of glass and thus are non-flammable and not susceptible to corrosion, ensuring low maintenance costs.

Building better rail networks
In addition to Hong Kong's rail network, rail operators around the world are beginning to adopt the new technology in their systems. The Beijing–Shanghai High-Speed Railway was the first to test the technologies in 2011. More recently, Transport New South Wales in Australia is planning to install the sensors on its first driverless trains.

As a result of cooperation with Dalian Jiaotong University in China, the PolyU-developed FBG sensors have been installed in high-speed inspection trains for monitoring the condition of rails. Another user, Southwest Jiaotong University, through its Railway Development Company, recently expanded its cooperation with PolyU's Base for Commercialization, Training and Research in Shenzhen. Having already collaborated

on the installation of FBG sensors in high-speed trains on the mainland, the two parties signed an agreement in June 2014 to further advance the development of the high-speed rail network and technology.

Fostering new perspectives in niche areas
Building on a solid foundation of research strengths established over the years and the synergies achieved through multi-disciplinary collaboration, the university has been developing novel innovations and raising its profile both regionally and internationally.

Developing sophisticated space tools – camera pointing system
PolyU has been continuously expanding its space engineering technologies on several fronts. Recent successes include a PolyU designed and built camera pointing system that was used to capture images of the Moon's surface by the Chinese lunar lander Chang'e-3, which landed on the Moon in December 2013. Jointly developed by PolyU researchers and the China Academy of Space Technology, it is the first Hong Kong-designed and -manufactured instrument to operate successfully on a celestial body. A lightweight motorized device, it moves the camera vertically and sideways so that it can capture images of the Moon and monitor the movement of the lander's rover. All pictures sent back by the lander from the surface of the moon were taken with the support of this device.

Wu Bo of the Department of Land Surveying and Geo-Informatics also contributed to the success of the mission, as his lunar mapping techniques were used

in the topographical analysis and selection of Chang'e-3's landing site.

A PolyU research team has also been invited to develop the surface sampling and packing system for the third phase of China's Lunar Exploration Programme, which is to be launched in 2017. The team is led by Yung Kai-leung of the Department of Industrial and Systems Engineering and with the support of Robert Tam and his teammates at the Industrial Centre, who will participate in the mission of bringing lunar regolith (loose rock and dust above a layer of bedrock) back to earth.

Experienced in international space exploration, the PolyU team has developed many flight-qualified space instruments over the years, including the Mars Rock Corer for the European Space Agency's Mars Express Mission in June 2003. It also developed the Soil Preparation System (SOPSYS), which was installed in the lander of the Sino–Russian Phobos-Grunt mission in November 2011. SOPSYS conducted *in situ* analysis of Phobos rocks.

At the 42nd International Exhibition of Inventions held in Geneva in April 2014, SOPSYS was awarded the Grand Prize and Gold Medal with Jury's Commendation.

A DECADE OF GROUNDBREAKING MEDICAL RESEARCH

The Institutes of Biomedical Sciences (IBS) is a research centre established to promote cooperation between basic research and clinical studies at Shanghai Medical College of Fudan University. Successively led by Fuchu He, Lin He, and Peng-Yuan Yang, IBS consists of three research units and a core facility conducting research in metabolic and molecular cell biology, epigenetic medicine, birth defects and cancer biology, proteomics and systems biology for medicine, structural biology and innovative drug discovery. Researchers at IBS are encouraged to collaborate with other departments of Fudan University as well as hospitals affiliated to the Medical College.

Metabolic changes caused by cancer

Metabolic alteration is a hallmark of cancer cells. The Laboratory of Molecular Cell Biology, co-founded and co-led by Xiong Yue and Kun-Liang Guan, and later joined by Qun-Ying Lei, Shi-Min Zhao, Dan Ye and Haixin Yuan has discovered that protein acetylation is an evolutionarily conserved post-transcriptional modification that regulates nearly every metabolic enzyme involved in diverse metabolic pathways. The team demonstrated that protein lysine acetylation plays a key role in regulating metabolic enzymes and thus in maintaining metabolite homeostasis. The team further revealed the link between dysregulation of metabolism and multiple types of human cancers. They found

that eight genes encoding for four metabolic enzymes are frequently mutated in cancers, thereby altering epigenetic regulation. The researchers are currently investigating acetylation-based regulation in other cellular pathways and the metabolic changes that contribute to tumorigenesis for cancer therapy.

Histone modification and DNA methylation

Histone modification and DNA methylation are critical mechanisms for epigenetically altering gene expression. The Epigenetics Research Center, co-founded and co-led by Yang Shi and Yujiang Geno Shi, and later joined by Yanhui Xu and Fei Lan, has investigated the dynamic regulation of DNA methylation by the DNMT and TET protein families, which write and erase DNA methylation in the mammalian genome, respectively. Significantly, they used a structural approach to uncover the substrate regulation mechanism of TET2 and reveal the mode of crosstalk between histone modification proteins and de novo DNA methyltransferases (DNMT3A). The researchers are also using integrated approaches to develop small molecule inhibitors with potential therapeutic applications.

Deciphering RNA and neuronal differentiation

The Alastair Murchie laboratory focuses on characterizing ligand-sensing RNAs and

the mechanisms and functions of RNA sensors in bacteria and eukaryotes. The laboratory has identified and characterized a new class of aminoglycoside-sensing RNA that regulates the expression of aminoglycoside antibiotic-resistance genes. This was the first report of antibiotic resistance controlled by an antibiotic-sensing RNA.

Wenyu Wen's laboratory focuses on the molecular basis of protein complexes that regulate neuronal differentiation and signalling. The laboratory has performed structural and functional studies on the dynamic visual scaffold INAD and found hints of a novel mechanism by which scaffold proteins actively influence signalling processes by modulating the conformation-coupled redox potential.

IBS has made great strides in improving our understanding of human health and disease and will continue to bridge the gap between basic research and clinical applications. All of the research teams at IBS continue to recruit new faculty members and postdoctoral researchers to contribute significantly to the health needs of China and the world.

Fudan University
www.ibsfudan.org/en_index.asp

▶ Brasilia. "Marina does not negotiate on principles," she says.

Others go a step farther. "I think she is a little too radical," says Pedro Alves Vieira, a geologist at the State University of Goiás and environment secretary for the municipality of Goiás. Although Silva has hewed to the centre during the campaign, Vieira fears that she may be too focused on the environment, too religious and too rigid for the presidency.

"I think it's time for a change."

But some of Silva's major stances have evolved. As environment minister, she opposed the introduction of genetically modified soya beans, but has since sought to mend fences with the powerful agribusiness industry — stressing that modern agriculture, practised legally, is not at odds with forest protection.

Her plan for sustainable development focuses on stemming Brazil's gradual shift toward fossil fuels: renewable energy fell from 95% of the nation's energy mix in the 1990s to 78% in 2013, according to her campaign. To reverse that trend, she is calling for the installation of solar panels on 1 million homes and the revitalization of the sugar-cane ethanol sector, which has struggled to compete with heavily subsidized petrol. She is also pushing for sustainable forest concessions geared toward biomass energy.

By contrast, Rousseff has put more emphasis on dozens of large-scale hydro-electric dams that would increase development pressure on the Amazon rainforest. "We need more energy, but Brazil's current strategy is really based on big infrastructure," says Mercedes Bustamante, an ecologist at the University of Brasilia who spent two years at the science ministry under Rousseff. "I think it's time for a change."

Science policy has received little attention in the presidential campaign, although Silva has called for federal and private investments in research and development to expand from 1.1% of Brazil's gross domestic product to 2%. The government's science spending has increased over the past decade, but researchers say that it has also been spread thin by new programmes.

Even if she takes the helm, Silva will have to tackle shifting political winds in Brasilia. In 2012, the nation's congress voted to weaken the country's forest law, and some lawmakers are now pushing to scale back protected lands to allow mining and energy development.

But Schwartzman warns against underestimating Silva's political skills should she win. "She and the others who worked with Chico Mendes in Acre really came through a crucible," he says. "In some ways, looking back on that experience, I'm not surprised that she is where she is now." ∎

The Chinese Academy of Sciences, which employs 60,000 people, opens its 2014 congress in Beijing.

POLICY

Chinese science gets mass transformation

Teamwork at centre of Chinese Academy of Sciences reform.

BY DAVID CYRANOSKI IN BEIJING

Change is coming to the institute that has been at the heart of China's scientific development since the communist state began. The Chinese Academy of Sciences (CAS) is making unprecedented structural reforms to foster collaboration and turbocharge research. Proponents say that the initiative will make China a world leader in areas from neuroscience to particle physics; detractors question whether modern China needs such a sprawling organization at all.

The CAS, headquartered in Beijing, employs 60,000 people, and has 104 research institutes and a budget of roughly 42 billion renminbi (US$6.8 billion). Just under half comes from the government, to cover overhead costs and salaries; the rest comes from competitive government grants. It is "the biggest reform in the academy's history", CAS president Chunli Bai told *Nature*'s news team in an interview on 22 September.

The initiative — which was designed by Bai — follows an appeal by Chinese president Xi Jinping in July 2013 for the CAS to become a world leader in science. Although China's spending on science now outpaces that of all countries except the United States, many agree that the country is not getting a good return on its investment, in terms of both basic-science breakthroughs and commercialization of those findings.

Some underperformance might result from low salaries for scientists — including those at the CAS, says Bai — who scramble to supplement their incomes with multiple grants. "Their research becomes very fragmented," he said.

Scientists also shy away from collaborations because co-authorship dilutes their achievements in the eyes of grant committees. At a conference in 2012, Bai discovered that researchers across 20 different CAS institutes were working on related projects, but often did not know each other — duplicating work and missing opportunities to share knowledge. He also notes that they had little interest in discussing applications with industry representatives.

To deal with these issues, Bai has grouped research into four categories, and promising CAS scientists are being selected to work together under them. They will see their salaries jump, relying on grants for only 30% of their income, rather than for 70%. The idea is to encourage scientists to collaborate on fewer, large problems, rather than to churn out marginal advances in disparate projects that can be used to seek multiple grants. "Teamwork is the key word," says Mu-Ming Poo, director of the CAS Institute of Neuroscience in Shanghai.

The first category is devoted to establishing centres of excellence focused on basic science,

and will target areas in which China has a chance to dominate. One team, for which Poo has already selected 40 researchers from 11 different CAS institutes, specializes in brain science — one goal is to produce a transgenic monkey model that could be useful for neurodegenerative-disease research.

Tandong Yao, director of the Institute of Tibetan Plateau Research in Beijing will head another centre focused on the seismically active plateau: his team wants to probe how the crust underneath deformed over time. Other centres will focus on particle physics, thorium-based molten-salt reactors, and quantum information. The plan is to take the count up to 20 centres by 2020.

A second category will target areas with underdeveloped commercial potential, including microsatellites, marine information technology and drug development. A third category will establish collaborations around large-scale facilities: a synchrotron and a protein-science centre, for example, both in Shanghai, will be linked to make it easier to carry out high-precision protein studies and support academic and industry researchers, including some from abroad.

The fourth category, still in its design phase, will be devoted to initiatives that assist local development and sustainability. Its first will probably focus on work by the Institute of Mountain Hazards and Environment in Chengdu, which could help communities to prepare for natural disasters such as landslides.

But not everyone agrees that the changes tackle the key problems at the CAS. One Chinese university scientist, who is familiar with the reforms and did not want to be identified, says that the source of the problems is that the CAS hires too many young scientists who then go through little or no review, and essentially receive tenure without having to prove themselves. Bai counters that the opportunities offered by his initiatives will "encourage the underperformers to be better". He says that he has steered away from a US-style system in which underperformers can be fired: "It is a reform with Chinese characteristics."

Cong Cao, a science-policy specialist at the University of Nottingham, UK, says that the CAS is too big and unwieldy. "There's no organization in the world with so many functions," he says. He says that universities and individual research institutes, which are increasingly competing with the CAS for leading scientists, can offer a better management environment for scientists. "There probably is no reason for the CAS in China right now."

But Bai uses the same trait to justify the academy's existence: he argues that it allows the CAS to make wide contributions to China's development. "You have to understand the history," he says. ∎

Global Ebola response kicks into gear at last

US and UN mobilizations are welcome but overdue.

BY DECLAN BUTLER

The international community is responding to the Ebola outbreak — at long last. On 18 September, the United Nations Security Council adopted a resolution declaring the outbreak a "threat to international peace and security"; two days before, the United States had said that it will send 3,000 military personnel to West Africa and spend US$750 million to support civilian efforts. Both moves are unprecedented for a disease outbreak and just what is needed, say experts. But there are fears that quelling the outbreak now will be more difficult than if it had been tackled earlier.

The Ebola outbreak was first detected in March, but only the humanitarian group Médecins Sans Frontières (MSF, or Doctors Without Borders) has so far deployed substantial relief in the affected region, which centres on Liberia, Sierra Leone and Guinea (for more, see nature.com/ebola). MSF has repeatedly demanded that the international community launch a major effort to tackle the outbreak.

The UN resolution calls on member states to respond urgently. As it was passed, UN secretary-general Ban Ki-moon launched the United Nations Mission for Ebola Emergency Response to deploy resources from UN agencies in coordination with efforts from member nations and donors. The mission's remit includes "stopping the outbreak, treating the infected, ensuring essential services, preserving stability and preventing further outbreaks," said Ban.

The United States plans to establish military-run command and control headquarters in Liberia and build 17 treatment centres of 100 beds each in the affected region. It hopes to recruit medical staff to run them, and train up to 500 health-care workers a week.

The public-health community tends to view military involvement "with suspicion and mistrust", says Adam Kamradt-Scott, a health-policy researcher at the University of Sydney in Australia. But he welcomes the US plan: "Military forces have unique skill sets that can assist civilian authorities." If the operation does help to counter Ebola, he adds, it could set a new precedent for responding to disease outbreaks of international concern.

Training forces in outbreak response will take time, warns Daniel Bausch, who studies infectious diseases at Tulane University in New Orleans, Louisiana, and is caring for people with Ebola in Sierra Leone. He adds that the situation is too dire for the United States alone to solve, and that other nations must step in, something that the UN resolution should catalyse. It was passed by unanimous vote and had a high number of member-state sponsors.

Jean-Clément Cabrol, director of operations for MSF in Geneva, Switzerland, says that both efforts are welcome, but that the situation has worsened in the past few weeks, so speed is key. Treatment centres in the three countries most affected are completely overwhelmed (see page 474). This is fuelling the spread of the disease, because centres are forced to send away infected people, who then contaminate others.

Conventional public-health measures are no longer working, says Bausch. It has become next to impossible to detect and isolate those infected, and to monitor all the people they have been in contact with. He recommends focusing surveillance on areas next to affected zones to prevent further geographical spread of Ebola. He adds that international efforts will be effective only if they include training for health-care staff, for example in biosafety. Money and supplies are welcome, but by themselves, they are not enough. ∎ **SEE EDITORIAL P.459**
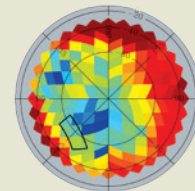
"Military forces have unique skill sets that can assist civilian authorities."

→ MORE ONLINE

TOP NEWS

● Population boom to last all century go.nature.com/xnsblj
● Male astronomers get bigger share of Hubble go.nature.com/v9e8o7

IMAGE OF THE WEEK

This map measures dust in the southern Galactic hemisphere. It suggests that a signal of gravitational waves from the Big Bang previously observed in the black rectangle could have been spurious.
go.nature.com/ifalmw

COMMENT

Jie Zhang (right), president of Shanghai Jiao Tong University awards an honorary doctorate to Peter Salovey, president of Yale University.

Chinese university reform in three steps

High-quality faculty, valued and rewarded, is the key to building a world-class research institution, says **Jie Zhang**.

China's economic growth is slowing after 35 years of rapid expansion. Sustainable development depends on converting that growth into innovation. So the Chinese government has substantially increased its investment in universities and research institutes. In 2012, for example, it spent more than 1 trillion renminbi (US$161 billion) on research and development and more than 700 billion renminbi on higher education.

As a result, research capacity and productivity have grown. Between 2005 and 2012, the number of full-time-equivalent researchers in China increased by 38% (to 314,000). Over the same period, the number of published research articles from Chinese higher-education institutions rose by 54% (to 1,117,742) and granted patents went up eightfold (to 66,755).

Yet the quality of research, as indicated by citations, lags behind, and technology transfer is sluggish. Ossified practices in

THE UNIVERSITY EXPERIMENT
A *Nature* special issue
nature.com/universities

evaluation and incentivization — such as rewarding publication quantity over quality — are holding Chinese universities back.

As president of a Chinese research university, I believe that building a high-quality faculty is the key to developing a world-class research university. In the past ten years, my institution, Shanghai Jiao Tong University (SJTU), has created a culture of innovation and boosted research capacity through reforms to career paths for existing and new faculty members.

China's leading universities are taking steps to enhance their research productivity. ▶

A robot developed by the School of Mechanical Engineering at Shanghai Jiao Tong University in China.

▶ Setting up new institutes, hiring high-profile overseas scholars at high salaries and introducing tenure systems are proven shortcuts to excellence. But importing a handful of 'star' professors is not sufficient to change university culture. Indeed, they can make local faculty members feel overlooked. Instead, fundamental reforms are needed.

My experience at SJTU presents a good worked example. We have implemented a university-wide incentive system to motivate all faculty, staff and students. The goal is to develop a high-quality faculty comparable to that of the best Western universities by 2020.

The changes began in 2007 with an analysis of future national and global challenges, open discussions among faculty and staff, and a feasibility review. Priorities emerged, including engaging creative minds, incubating innovations and bridging the sciences and humanities to best serve the nation and the world. To maintain faculty support, fundamental reforms were adopted in stages over several years.

First: recruit and mentor junior staff. World-class scientists were hired to build research groups focused on cutting-edge science and engineering problems and to set a high bar for academic performance. Junior faculty members were recruited competitively in the international job market. Generous start-up funds supported the groups for the first few years, until they could attract funds from Chinese government agencies or industries, start to produce results and establish their reputations.

A tenure system for new faculty members began in 2007. A six-year tenure track was set up for new junior faculty based on midterm and final evaluation by an international committee.

In 2008, we launched a mentoring system for junior faculty under 35 years old. And since then, 1,251 junior faculty members have won extra support for their research, housing and living costs from a 150-million-renminbi endowment.

Second: three career tracks for faculty. Promotion and salary are assessed every three years against performance indicators developed by each school or department. Three career tracks — teaching, research and tenure — were created in 2010 for existing faculty. The teaching track has no research expectation; scientists with no teaching obligation must make up part of their income from competitive research grants. Tenure-track professors teach and do research and are evaluated according to combined criteria.

Each track has similar starting salary scales. Faculty members chose their own track, with the option to switch later if peer evaluations allow. All faculty may apply for transfer to the tenure-track system as standards rise; the best will be encouraged to do so. The average salary for faculty has increased by 60% in the past four years, and will rise further in the next four years.

The university decentralized its governance to give schools and departments more autonomy to recruit, develop and evaluate their staff. Budget reforms have given each school or department direct access to resources.

Third: one merged tenure system. The infusion of fresh academic blood gave the university a split nature: international scholars on the Western tenure path; existing faculty following the three tracks. In 2013, six pilot schools or departments began to merge the two paths into a single tenure system similar to that of North American universities. Encouraging

results have already been seen in physics, mathematics, mechanical engineering, biomedical engineering, law and management.

On the basis of these experiences, the entire university will switch to the tenure system between 2015 and 2018. Although it is impossible to suddenly double or triple the salary scale for all 3,000 faculty members, mechanisms for paying competitive salaries within the tenure system are being built, such as setting up chairs and fellowships from a 500-million-renminbi endowment.

Faculty members who do not qualify for the tenure track can choose to leave or remain in the contract system until it is phased out by 2018. It is important that faculty who have served the university well but cannot meet the new standards be treated fairly and with respect and provided with a channel to continue to serve the institution.

PROVEN RESULTS

Today, SJTU is well on its way to being one of the leading research universities in the world. Since 2007, 450 world-class professors and top-tier young faculty members of international standing have joined the university, and more than 250 existing faculty members have transferred to the tenure-track system. Eighty-five per cent of faculty hold PhDs (up from 50% in 2006).

The university's annual revenue has more than doubled since 2007, to more than 7 billion renminbi; competitive research income has tripled to more than 2.5 billion renminbi. The number of disciplines offered at SJTU that are ranked in the global top 1% of Thomson Reuters' Essential Science Indicator (ESI) jumped from 5 in 2007 to 16 in 2014. Social sciences joined engineering, and natural, life and medical sciences in the top 1%, making the university more comprehensive.

In terms of number of papers published, the university has ranked second globally in engineering since 2007. ESI ranked natural sciences 40 in 2013, up from 57 in 2007. Life sciences and agriculture rose to 43 from 136, and medicine is now ranked 27, up from 166. Citations are up and more discoveries are being patented.

The reforms at SJTU have promoted a shift in educational emphasis. We are moving away from knowledge transfer to knowledge creation and from instruction-centred teaching to student-centred learning. Our philosophy has changed to nurturing students to be engaged, competent global citizens. A culture that values and rewards innovation has successfully taken root. ∎

Jie Zhang *is the president of Shanghai Jiao Tong University in Shanghai, China. He is an academician of the Chinese Academy of Sciences, and a foreign associate of US National Academy of Sciences.*
e-mail: jzhang1@sjtu.edu.cn

IMAGINECHINA/REX

BIOMEDICINE

China opens translational medicine centre in Shanghai

First of five linked institutes aims to capitalize on basic-research investments.

BY DAVID CYRANOSKI

Over the past decade, China has poured money into research, especially in the biomedical sciences. But as the nation's health-care costs have risen in the past few years, critics have argued that the investment has not paid off. A group of researchers and government officials now hopes to improve those returns with the official opening this month of the National Centre for Translational Medicine in Shanghai. The 1-billion-renminbi (US$163-million) centre, slated to complete construction in 2017, is the first of five institutions meant to bridge the gap between basic research and clinical application by putting researchers, doctors and patients under one roof.

China's biologists have made impressive progress in fields such as genome sequencing and protein-structure analysis, but have produced little in terms of drugs or other medical products. "Bloggers and others are always complaining that China is just burning money," says Xiao-Fan Wang, a cancer researcher at Duke University in Durham, North Carolina, and one of the 21 people on the centre's international consulting committee. Some even ask why China, which has made progress in industry largely by copying other countries, has not succeeded by following the same strategy in the biomedical sciences, says Wang.

Wang says that doctors in China are overworked, often racing through whole days of back-to-back patient visits and procedures. And yet, because most Chinese hospitals are part of universities, the doctors must publish to get promoted. In this competitive environment, they often refuse to share data, but are rarely able to do thorough research themselves. "You can't expect them to do eight hours of surgery and then jump into a lab coat," says Wang. The Shanghai centre will change that, he says, by giving clinicians time to focus on research questions.

Saijuan Chen, the centre's director and a geneticist at Shanghai Jiaotong University, says that the centre began interviews this month to recruit some 50 principal investigators and 12 scientists to direct research in their disciplines. She says that the institution's focus will be on developing treatments for heart disease, stroke, metabolic diseases and cancer.

The centre's international consulting committee will help to recruit top talent in these

A major push is under way in China to develop new drugs and surgical procedures.

fields. Tak Mak, a committee member and an immunologist at the University of Toronto in Canada, says that the strategy shows a commitment to hiring on the basis of expertise rather than connections. It is "an effort to break out of the old system — which is probably the system in more places than you'd care to know", he says.

FINE BALANCE

One challenge will be attracting Chinese biomedical researchers who are working abroad. Wang says that China has been able to lure basic-research scientists back to the country with competitive pay. But clinicians in China receive a fraction of the salaries they can get in the United States.

Managing the centre may also be difficult, says committee member Sujuan Ba, chief operating officer of the US National Foundation for Cancer Research in Bethesda, Maryland. The centre's governing council comprises members from ten government agencies and institutions. "That shows the wide scope of support from China, but it is going to be a huge challenge for the leadership team to balance and to manage each council member's requisites," she says. "It is very important for the centre to steer away from bureaucratic red tape and stay focused on its mission and long-term vision of conducting high-impact translational research."

The 54,000-square-metre site will have 300 beds for patients and study volunteers.

It will also run a biobank that will collect hundreds of thousands of patient samples, and an 'omics' centre that will conduct high-throughput genome analyses and gather data on proteins and metabolic products.

Haematologist Zhu Chen of Shanghai Jiao Tong University, who is China's former health minister and chairman of the centre's scientific advisory board, hopes that the centre will emulate the spirit of St Jude Children's Research Hospital in Memphis, Tennessee, with its close relationship between clinicians and basic researchers. He also emphasizes the need to make therapeutic trials free: in China, participants are often charged for treatment. Zhu and Saijuan Chen, who are married, led one of the few successful translational research projects in China — treatment of a form of leukaemia using retinoic acid and arsenic trioxide.

The Shanghai centre will eventually have four sister institutions: a geriatrics centre at the People's Liberation Army general hospital in Beijing, a centre for rare and refractory diseases at the Peking Union Medical College in Beijing, a molecular-medicine research centre at the Fourth Military Medical University in Xian and a regenerative-medicine centre at the West China Hospital in Chengdu.

"These centres are at the historic moment to make a huge impact for China's drug development," says Ba. "We should be able to see the signs of success in five years, if not earlier." ■